Comparative Cladistics: Fossils, Morphological Data Partitions and Lost Branches in the Fossil Tree of Life

Mounce, Ross

Award date:
2013

Awarding institution:
University of Bath

Comparative Cladistics:

Fossils, Morphological Data Partitions

and Lost Branches in the Fossil Tree of Life

Ross Callum Parker Mounce

http://orcid.org/0000-0002-3520-2046

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

October 2013

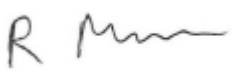
Signature: 

Table of Contents

Acknowledgements

I would like to thank my supervisor, Matthew Wills for putting up with me all this time. I have been lucky to have such accommodating and understanding support. I also must thank my lab mates Martin Hughes, Anne O'Connor, Sylvain Gerber, Katie Davis, Rob Sansom and everyone else in the Biodiversity Lab at the University of Bath – we had some great times and some brilliant times together. Sincere thanks also to the University of Oslo Bioportal computing cluster for providing me free cloud computation for my work.

Many people have helped spur my imagination along the way with ideas for different chapters of this thesis. For this I would like to thank Ward Wheeler, Pablo Goloboff, Mark Siddall, Dan Janies, Steve Farris and the generous financial support of the Willi Hennig Society. I want to thank all those in the palaeontology community who have shared their published data with me, particularly Graeme Lloyd for his stirling work in making dinosaurian data available – I hope I have done something interesting with the data I have used and opened eyes to new possibilities. I also want to thank all those in the open science community – Peter Murray-Rust, Todd Vision, Heather Piwowar, Mark Hahnel, Martin Fenner, Geoffrey Boulton, Jenny Molloy and so many more I've had the pleasure of meeting in person. The energy and enthusiasm I drew from countless online discussions on Facebook, Google+ and Twitter was truly inspirational.

For facilitating greater access to scientific literature I must heartily thank the Natural History Museum, London library and archives, the #icanhazpdf community on Twitter, Wikipaleo on Facebook, References Wanted on FriendFeed, Library.nu, and SciHub. Without these additional literature access facilitators I would not have been able to read half the sources I cite in this thesis.

I must thank my wife Avril for her patience with me especially during the write-up phase, for allowing me to go away to all these amazing conferences abroad, and for tolerating all those long nights into mid-morning when I was tapping away on my noisy keyboard. Finally, I thank my family: Richard, Rosemary & Tara for repeatedly encouraging me to finish my thesis – I got there in the end!

Publications

- Mounce, R. C. P. and Wills, M.A. 2011 Phylogenetic position of Diania challenged. *Nature* 476:E1.
- Mounce, R., Butler, A., Davis, K., Dunhill, A., Garwood, R., Lamsdell, J., Legg, D., Lloyd, G., Pittman, M., Warnock, R., and Wolfe, J. 2011. An open letter in support of palaeontological digital data archiving. http://supportpalaeodataarchiving.co.uk/
  (featured in *Nature News*: 'Fossil data enter the web period' doi:10.1038/472150a)
- Stoltzfus, A., O'Meara, B., Whitacre, J., Mounce, R., Gillespie, E., Kumar, S., Rosauer, D., and Vos, R. 2012. Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes* Vol. 5, 1-15.
- Mounce R. 2012. Twitter for Systematists. *The Systematist* Vol. 34, 14-15.
- Mounce, R. 2012. Life as a palaeontologist: Academia, the internet and creative commons. *Palaeontology Online* Vol. 2, 1-10.
- Mounce, R. 2013. Open access and altmetrics: distinct but complementary. *Bulletin of the American Society for Information Science and Technology* Vol. 39, 14-17.
- Mounce, R. and Janies, D. 2013. Synergistic sharing of data and tools to enable team science. In: J. Zander and P. J. Mosterman (eds.) Computation for Humanity: Information Technology to Advance Society. CRC Press, 437-447.
- Poisot, T., Mounce, R. and Gravel D. 2013. Moving toward a sustainable ecological science: don't let data go to waste! *Ideas in Ecology & Evolution* Vol. 6, 11-19.
- Mounce, R. and Wills, M. A. (in revision) Congruence of Cranial and Postcranial Characters in Vertebrate Phylogeny

Media

Radio: BBC Radio 3 Night Waves http://bit.ly/NightWaves [live, 2012-10-02]
Voice of Russia radio http://bit.ly/VoRussia [recorded as live, 2012-08-17]

Press mentions in: Publishing: Open to possibilities [Nature Jobs, 2013-03-27]
Researchers opt to limit uses of open-access publications [Nature News, 2013-02-06]
Text-mining spat heats up [Nature News, 2013-03-20]

List of Abbreviations

AAAS	American Association for the Advancement of Science
AgD	Agreement Subtree Distance
AMNH	American Museum of Natural History
ASIH	American Society of Ichthyologists and Herpetologists
ASPT	American Society of Plant Taxonomists
*BEAST	Bayesian Evolutionary Analysis by Sampling Trees (software)
BMC	BioMed Central
CFI	Consensus Fork Index
CI	Consistency Index (ensemble)
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CUP	Cambridge University Press
DNA	Deoxyribonucleic acid
GHz	Gigahertz
GIC	Group inclusion characters
GNU	GNU's Not Unix!
GS	Google Scholar
GUI	Graphical User Interface
HDD	Hard Disk Drive
HER	Homoplasy Excess Ratio
HERM	Homoplasy Excess Ratio Maximum
HTML	HyperText Markup Language
I/O	Input/Output
ILD	Incongruence Length Difference
IRD	Incongruence Relationship Difference
KH	Kishino-Hasegawa
MAS	Microsoft Academic Search
MAW	Dr Matthew A. Wills (supervisor)
MEANNS	Mean Optimal Tree Length Inferred after Matrix Randomizations
MEGA	Molecular Evolutionary Genetics Analysis (software)
MF	Mickevich-Farris
MHER	Modified Homoplasy Excess Ratio
MINL	Minimum Possible Length

MPT	Most Parsimonious Tree
MS	Mendeley Search
MWU	Mann-Whitney U
NNID	Nearest Neighbour Interchange Distance
NONA	No Name (software)
NPG	Nature Publishing Group
OA	Open Access
OSI	Open Source Initiative
OUP	Oxford University Press
PAUP*	Phylogenetic Analysis Using Parsimony *and other methods (software)
PDF	Portable Document Format
PHYLIP	PHYLogeny Inference Package (software)
PICA	Phylogenetic Inference by Compatibility Analysis (software)
PLD / PD	Path Length Difference / Path Difference
PLOS	Public Library of Science
PMC	PubMed Central
PTP	Permutation Tail Probability
QD	Quartets Distance
RAS	Random Addition Sequences
RAxML	Randomized A(x)ccelerated Maximum Likelihood (software)
REHER	Relative Expected Homoplasy Excess Ratio
RF	Robinson-Foulds
RI	Retention Index
SSD	Solid State Drive
STEM	Species Tree Estmation using Maximum Likelihood (software)
TAXEQ3	Safe TAXonomic reduction using taxonomic Equivalence (software)
TBR	Tree Bisection and Reconnection
TILD	Topological Incongruence Length Difference
TNT	Tree analysis using New Technology (software)
tsv	Tab-separated values
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
WoK	Web of Knowledge
XML	Extensible Markup Language

List of Figures

List of Tables

Summary

In this thesis I attempt to gather together a wide range of cladistic analysis of fossil and extant taxa representing a diverse array of phylogenetic groups. I use this data to quantitatively compare the effect of fossil taxa relative to extant taxa in terms of support for relationships, number of most parsimonious trees (MPTs) and leaf stability. In line with previous studies I find that the effects of fossil taxa are seldom different to extant taxa – although I highlight some interesting exceptions. I also use this data to compare the phylogenetic signal *within* vertebrate morphological data sets, by choosing to compare cranial data to postcranial data.

Comparisons between molecular data and morphological data have been previously well explored, as have signals between different molecular loci. But comparative signal *within* morphological data sets is much less commonly characterized and certainly not across a wide array of clades. With this analysis I show that there are many studies in which the evidence provided by cranial data appears to be be significantly incongruent with the postcranial data – more than one would expect to see just by the effect of chance and noise alone.

I devise and implement a modification to a rarely used measure of homoplasy that will hopefully encourage its wider usage. Previously it had some undesirable bias associated with the distribution of missing data in a dataset, but my modification controls for this. I also take an in-depth and extensive review of the ILD test, noting it is often misused or reported poorly, even in recent studies.

Finally, in attempting to collect data and metadata on a large scale, I uncovered inefficiencies in the research publication system that obstruct re-use of data and scientific progress. I highlight the importance of replication and reproducibility – even simple re-analysis of high profile papers can turn up some very different results. Data is highly valuable and thus it must be retained and made available for further re-use to maximize the overall return on research investment.

Chapter 1: Introduction

Fossils provide a special 'window' through which we can glimpse the breadth and diversity of past morphological forms of life. More than 99% of all species that have ever existed are extinct (Novacek & Wheeler 1992; Nee & May 1997). Thus if we are to truly understand evolution we need to include extinct as well as extant forms. Through this window we have observed countless remarkable specimens of past organisms that can defy at-a-glance placement within our established schemes of classification and phylogeny for known lifeforms (e.g. *Hallucigenia* that was first described by Conway Morris [1977] in phylum “unknown”; see also *Anomalocaris*, Fig. 1). To better understand both what fossil specimens are, and to infer their relationships to other extant and extinct species, specimens are frequently compared on the basis of their morphological characteristics. These characteristics are coded into a matrix which can then be used to compare the morphological character scorings of similar organisms using numerical analyses to create a testable hypothesis of their relations – frequently expressed as a dendrogram that represents an estimate of the relations of the organisms studied.

Figure 1.1 An example of a remarkable fossil form. *Anomalocaris* from the Mt Stephen Trilobite Beds, Middle Cambrian, near Field, British Colombia, Canada. Credit: photo by Mark A. Wilson / Public Domain.

1.1 Overview of Inferring Evolutionary Relationships in Palaeontology

"There are parts of the palaeontological community where cladistic drums do not reach; there are nooks and crannies, fault lines, sink holes and caves where the words synapomorphy, paraphyly and parsimony have never been heard" Peter Forey recounting the words of his Editor (Palaeontology Newsletter), 2005

The history of classification in biological systematics can be traced back millennia to early scholars such as Aristotle, who produced one of the first recorded biological classifications. In his works such as *Historia Animalium* and *De Historia Plantarum* he grouped organisms on the basis of similarity. This system was not an *evolutionary* classification but it was nevertheless a noteworthy contribution. Aristotle also saw partial similarity between fossil shells and their modern counterparts but he misattributed their existence and formation as 'dry vapour exhalation' (Eichholz, 1949; Mayor, 2000). In the 18th century Linnaeus furthered the cause with his system of Linnaean taxonomy, separating nomenclature from description – a system which in essence we still use today. Linnaeus recognised three kingdoms: animal, vegetable and mineral. Although more widely-known for neontological contributions, Linnaeus also made significant contributions to palaeontology – some fossil corals still bear the same names as when he first described them after fieldwork in Gotland and elsewhere (Linnaeus, 1745; Linnaeus, 1758).

Post-Darwin, the history of classification and evolutionary relationships in palaeontology was spurred-on by George Gaylord Simpson who unlike most of his palaeontological colleagues (Olson, 1991) recognized the importance of genetics in evolutionary studies – in support of the 'modern evolutionary synthesis'. Papers such as 'Patterns of phylectic evolution' (Simpson, 1937) and his opus *Tempo and Mode in Evolution* (Simpson, 1944) used quantitative statistical reasoning to support his arguments of fossil mammal relationships and species delimitation. His work placed palaeontology as a more objective science because of this quantitative, statistical reasoning. In the 1960's, two distinct schools of quantitative methods for inferring evolutionary relationships arose – that of numerical taxonomy and that of phylogenetic systematics (cladistics). Numerical taxonomy again worked on the basis of just similarity. Sokal & Sneath (1963) in their book *Principles of Numerical Taxonomy* define numerical taxonomy as "the grouping by numerical methods of taxonomic units into taxa on the basis of their character states" (p4). Their explicit preference for similarity as the optimal criterion is made clear as early as page 6

where they write “Similarities among taxonomic entities can be represented geometrically by points in a space... The distances between the points can be regarded as taxonomic distances”. Today, few systematists, palaeontologists or otherwise use similarity as an optimality criterion for inferring phylogeny. Instead, the other major branch, phylogenetic systematics (cladistics) as it was called then and sometimes still is now, dominates as the preferred method with which to infer relationships.

These cladistic methods that we now use to infer phylogeny were first devised by a German neontologist; Willi Hennig (1950) who studied true flies (Diptera). But it wasn't until the publication of an English translation of a modified version of this book (Hennig, 1966), and various popularisers (e.g. Brundin, 1966), that his methods gained widespread recognition.

For a time after this, most of palaeontology seemed slow and even reticent to adopt these new methods on fossils. Looking back on it, commentators such as Siddall (1998) explained this resistance as attributable to “the follies of ancestor worship” - using the temporal sequence of fossils to *assert* ancestor-desendent relationships between fossils, rather than *testing* relationships of common ancestry. Hennig's 'phylogenetic systematics' (which we now refer to as cladistics) was met with “distrust” by some palaeontologists and viewed as a threat to the primacy of palaeontology's role in tracing phylogenetic pathways (Forey 2004). Prior to the Hennigian revolution, palaeontological data had been considered “both necessary and sufficient” for phylogeny reconstruction (Forey 2004), cladistics challenged this. A notable early exception to the distrust was a chapter by Schaeffer *et al.* (1972); three palaeontologists that clearly advocated the use of cladistic principles in palaeontology:

> “...we agree that (1) the degree of relationship (as defined by recency of common ancestry) should be determined initially on the basis of morphology alone...”

Even then though, there was still a hesitance as to whether one could or should mix recent and fossil taxa in the same analysis. Among those who helped make it clear that recent & fossil taxa *could* in fact be used in the same analysis together was Farris (1976):

> “Fossil species – when they are sufficiently well known to be classified at all – should therefore be treated exactly as are recent species in a phylogenetic classification”.

Patterson's (1981) review of the significance of fossils in determining evolutionary

relationships represents a turning point in palaeontological thinking:

> "before the development of the cladistic method paleontology was a hindrance rather than a help, stifling progress towards the goal of determining evolutionary relationships"

But in his review Patterson (1981) also peculiarly (considering he was a palaeontologist) belittles the role of fossils with his oft-cited conclusion that it is rare or unknown for fossils to overthrow theories of relationship primarily based on Recent forms (p.219). This has been convincingly refuted many times since (e.g. Gauthier *et al.* 1988; Donoghue et al 1989; Eernisse and Kluge 1993; Cobbett et al 2007). Patterson (1981) also noted that the incompleteness of fossils specimens relative to Recent specimens makes them inherently less informative (but as was shown later it is not what is missing that counts *per se* but instead what information is present that most affects phylogenetic inferences e.g. Huelsenbeck 1991; Wiens 2003a,b; Edgecombe 2010).

The schism between palaeontological and neontological approaches to phylogenetic reconstruction was so fascinating that it even drew the attention of philosophers of science. Grantham (2004) expounded at great length on this matter evaluating a number of hypotheses as to the root cause of the integration difficulty, and concludes that it is a relatively unique phenomenon relative to other fields. For a more complete history in the fuller context of systematics read Forey (2004).

1.2 Why not just use molecular data?

Morphology for most fossil specimens is all the evidence we have. Whilst for Recent organisms we nearly always additionally or solely use molecular sequence evidence, this cannot be obtained from most fossils for a variety of reasons (aside from issues of 'difficulty' or 'expense'). In the field of ancient DNA studies, the oldest fossil from which a genome has been extracted is a ~700,000 year old horse (Orlando *et al.*, 2013). For the age of the oldest successfully extracted DNA fragments there is great controversy over the repeatability and validity of claims. Cano et al (1993) reported extracting DNA from an amber-entombed weevil that is 120-135 million-years-old. Many others have also claimed DNA extraction from geologically ancient fossil specimens (reviewed in Hebsgaard et al 2005). Yet subsequent attempts to repeat these feats have cast significant doubt on the validity of these claims (e.g. Austin et al 1997; Gutierrez & Marin 1998). Moreover, well-parameterized models estimate that the half-life of mitochondrial DNA (typically better

preserved than nuclear DNA because there is ~1,000 times more of it) is 521 years and thus even in optimal preservational conditions no mitochondrial sequence of greater than one base pair should remain after 6.8 million years (Allentoft et al 2012). Thus for fossils older than this, hard physio-biogeochemical constraints mean it is likely we will never be able to extract geologically ancient DNA samples from most fossils and hence must continue to use morphology-based methods of phylogenetic inference for the forseeable future with these taxa.

1.3 Is Morphology Actually Useful for Reconstructing Phylogeny?

In much of my introduction so far I have not addressed the very use of morphology itself (whether neontological or palaeontological) in reconstructing phylogeny. The use of morphology to infer phylogeny is not without its critics. Papers such as Scotland *et al.* (2003) have argued that fewer (but more rigorous) morphological characters should be used for phylogenetic inference. One of the points that Scotland et al (2003; hereinafter SEA) make is that the same observed morphological characters states can often be coded in very different ways. The choice of character coding method for most morphological character states is thus subjective and the use of these different methods alone can cause real differences in the resulting phylogeny. This is not such a problem for molecular sequence data where are mostly always just 4 states, but the inference of alignment gaps and their treatment somewhat complicates this. Another point made by SEA was that the conceptualization of morphological characters themselves is also subjective.

“Different workers will perceive and define characters in different ways” (Smith 1994, p34)

Simply put – using DNA sequence data is more objective than using morphological data because DNA offers large numbers of relatively unambiguous characters and character states. SEA concludes that:

> “morphology is being superseded by DNA data for phylogenetic studies because much of the useful morphological diversity has already been scrutinized... We disagree that morphology offers any hope for the future to resolve phylogeny at lower or higher taxonomic levels”

Needless to say, this controversial paper provoked some very direct responses (Jenner

2004a; Wiens 2004; Smith & Turner 2005). Jenner (2004a) dismisses SEA's points about the subjectivity of morphological homology assessment and character coding as a “straw man” argument. In contrast Jenner (2004a) instead predicts that the application of new analytical techniques will help further and improve, refine and expand our assessments of morphology for use in phylogenetic reconstruction. The application of new 3-dimensional imaging methods to help assess morphology certainly seems to support this point (e.g. Ragsdale and Baldwin, 2010). Another important counter-point from Jenner (2004a, p 337) is that there is no evidence to suggest that morphology *generally* performs more poorly than molecular data in fair comparisons. Jenner (2004a) cites examples where different molecular analyses conflict with each other, and where morphology appears to be more reliably than molecular data. Clearly there is no *a priori* general rule as to what type of data is 'better' for phylogenetic accuracy.

Wiens (2004) makes some subtly different points to Jenner (2004a), expanding upon the importance of fossils and hence the necessity of morphology. Wiens (2004) conveys that fossils help not just in determining phylogenetic relationships, but also in our understanding of the timing and rate of evolutionary processes – such uses of 'fossil calibration' points (reviewed in Donoghue and Benton 2007) are commonplace now. Wiens (2004) points out that SEA's argument against morphology because of its frequently incomplete nature is also irrelevant. Previous studies by Wiens (2003a,b) show that highly incomplete taxa can be placed with 100% 'accuracy' relative to simulated data, and that incompleteness may limit but does not outright prevent such incomplete taxa overturning the relationships of more complete taxa. Wiens (2004) also points out that for many taxa we have very few precious specimens for them, and that we cannot extract DNA from these owing to their rarity and sometimes the way in which they have been preserved – these also may require morphology-based assessment to determine phylogenetic placement. Wiens (2004) presents morphological data as a best “reality-check” for molecular studies and brushes away the criticisms presented by SEA, in that although some are problematic, the benefits of using morphology should outweigh the negatives. Wiens (2004) rejects SEA's claim that all the 'good' morphological characters have been found already and that there are few good ones left to add based on his own experience and publications. I must say that at a glance of any volume of *Journal of Vertebrate Palaeontology* you will clearly find studies that are finding new morphological characters, even in extremely well studied groups. High profile examples that have massively

expanded the suite of morphological characters for well studied groups include (squamata – Conrad 2008; post-Paleozoic echinoids – Kroh & Smith 2010; arthropods - Legg *et al.* 2012; placental mammals - O'Leary et al 2013).

Wiens (2004) repeatedly points to his study of congruence (Poe & Wiens 2000) which provides empirical evidence against some of SEA's points. Wiens (2004) conclusion is that instead of abandoning morphological phylogenetics we should instead work on extending and refining our morphological characters and characters states to address some of the problems that SEA point out.

The last of the direct critiques of SEA was published a little later (Smith & Turner, 2005). They claim that SEA misinterpreted some of the previous studies that SEA cited in support of their claims. In particular Smith & Turner (2005) rightly point out that Hillis (1996; 1998) used models of evolution and molecular data and that the conclusions from these analyses may not necessarily be transferable to morphological data which typically has very different properties. As Smith & Turner (2005) are both qualified systematic palaeontologists themselves they are well placed to rebutt SEA's assertion that most new morphological characters will be more homoplastic. They rightly point out that SEA provide no evidence for this, and helpfully suggest ways in which this could be formally tested. Smith & Turner (2005) similarly uphold the importance of fossils and morphology in phylogenetic analyses.

SEA's paper is far from the only one to criticize the very use of morphology but it is a rather obvious one on which to focus because it draws out many of the arguments for and against. Alternative philosophical approaches such as that of consilience (Wilson, 1998; Pisani, 2002) also support the use of both molecular and morphological data. I hope this short synopsis demonstrates that although questioned, the exclusion of morphology and fossils from phylogenetic analysis “is neither theoretically nor empirically defensible” (Edgecombe, 2010).

1.4 Aims of this thesis

This thesis is first and foremost, a synthesis of morphology-based phylogenetic literature. In each of these chapters I perform various comparative cladistic analyses with appropriate statistical power with which to test hypotheses that one couldn't otherwise attempt to answer with just a handful of data sets. The discovery, sampling, standardisation and assembly of evidence of these chapters is more rigorous and systematic than many similar comparative cladistic analyses that have been attempted before.

Chapter 2 is an examination of the congruence of phylogenetic signal *within* vertebrate morphological data sets. I compare the congruence of signal between cranial and postcranial partitions of data sets to answer questions over the levels of homoplasy in each and the significance of difference between the two if any. I also explore the performance of a new statistical test of congruence of relationships inferred from data, as proposed by MAW called the Incongruence Relationship Difference (IRD) test.

Chapter 3 is a re-examination of the impact of fossil taxa in mixed phylogenetic analyses that include both extinct and extant taxa, using newer, larger data matrices. In this chapter I re-implement pre-existing methods in a new more computationally-efficient pipeline that should encourage other systematists to use these methods to explore their own data.

Chapter 4 is a critical systematic review of the usage of the ILD test in the recent literature. In this chapter I observe a number of worrying trends in the usage of the ILD test that are inappropriate and unsupported by published evidence. I hypothesise how these usages took hold in the literature and conclude with clear guidelines at the end which should hopefully ameliorate any confusion.

Chapter 5 is the introduction of a modification to Archie's (1989) Homoplasy Excess Ratio that improves the statistic when in the presence of significant amounts of non-randomly distributed missing data (as it typical of many palaeomorphological data sets). I demonstrate the improvement to estimates of homoplasy that this modification gives, and I implement it with a script in TNT so that other investigators can also use it.

Chapter 6 is a critical examination of the fragmentation, discoverability and accessibility of phylogenetic knowledge in the modern age. I compare traditional literature search engines

such as Web of Knowledge that only search titles, abstracts and keywords, with local desktop full-text search methods over tens of thousands of papers. The findings in this chapter raise significant questions as to our ability to effectively synthesise data from thousands of papers

Chapter 7 provides my overall conclusions on what I have found with respect to the importance of fossils in phylogenetic reconstruction. But in this chapter I also take the opportunity to reflect the on some emergent themes in my thesis, namely; data availability, analysis replicatability, and how the way in which we choose to publish our research affects our ability to re-use data.

Chapter 2:

The Congruence of Cranial & Postcranial Characters in Vertebrate Phylogeny

Based on a manuscript co-authored with Matthew A. Wills

2.1 Abstract

Morphological data matrices frequently contain significantly more characters from some anatomical regions than others. Such preferential sampling is seldom considered problematic because homogeneity of phylogenetic signal across anatomical regions is almost invariably assumed. For this reason, signals within logical or anatomical partitions of morphological data sets are rarely compared. In vertebrate systematics, the cranium has often been afforded particular focus; either because it is believed to yield characters containing less homoplasy, or because morphological variation therein is more readily atomized. An analysis of 62 vertebrate data sets published between 2000 and 2010 confirms that characters of the cranium account for the significant majority, but finds equivocal evidence that they contain less homoplasy. We caution that neither partition ensemble consistency indices (CI) for partitions, nor mean per character consistency indices (ci) within partitions should be interpreted uncritically. Surprisingly, partition homogeneity (ILD) tests of the signal in cranial and postcranial characters reveals significant incongruence in a large minority of cases. Similarly, the trees inferred from the partitions are more different (Robinson Foulds and maximum agreement subtree distances) than expected about one time out of three (new randomization tests are proposed). This may reflect different selective pressures in particular body regions, allied with different localized patterns of homoplasy. In many cases, therefore, concentrating upon cranial characters at the expense of others (or *vice versa*) is quite likely to yield a phylogeny significantly different from that which would be obtained from a more holistic approach. We show that the broadest possible sampling of characters in the total evidence analysis of all aspects of morphology is most optimal, and caution against the assumption of signal homogeneity across all body regions.

2.2 Introduction

Phylogenies are typically inferred from morphological data by applying maximum parsimony to all coded characters. While studies often focus upon characters of particular types or from particular organ systems, it is rare that trees are inferred explicitly from subsets of these data (but see O'Leary *et al.*, 2003, Poyato-Ariza, 2003, Farke *et al.*, 2011), or that the signals from non-overlapping data sets (e.g., osteology and musculature) are compared. The practice of what is effectively a default total evidence analysis for morphology contrasts with the more qualified approach often adopted with molecular data. This is partly because sequence data are readily partitioned into logically distinct classes and sub-classes: nuclear genes, mitchondrial genes (plastid genes in plants), coding and non-coding sequences as well as codon positions (Bull *et al.*, 1993). It is widely understood that the signals from different types of molecular data (or indeed, from any two loci) can conflict (Felsenstein, 1988; Pamilo & Nei, 1988; Maddison, 1997; Nichols, 2001). This has been referred to as the 'gene tree discordance' problem (Degnan & Rosenberg, 2009), although we note that such discordance is not *necessarily* a result of different gene histories. In trees derived from modest numbers of molecular markers, signals may be variously tested for homogeneity prior to combination. At the other extreme, where the products of next-generation sequencing are assembled for phylogenomic analysis, it is common to apply filters of varying complexity to ensure a signal of some specified quality or homogeneity (Leigh *et al.*, 2011; von Reumont *et al.*, 2012).

Incongruence between partitions of molecular data can be attributed to a variety of causes: non-vertical inheritance affecting one or both partitions (e.g., gene loss/duplication, horizontal transfer, hybridisation or recombination), deep genetic divergence that does not reflect species phylogeny (also known as 'hemiplasy' or 'incomplete lineage sorting'; Avise & Robinson, 2008; Degnan & Rosenberg, 2009), or to a difference in the rate of evolution between partitions (Planet, 2006). Incongruence is also commonly observed between morphological and molecular data partitions (e.g., Mickevich & Farris, 1981; Bremer, 1996; Poe, 1996; Baker *et al.,* 1998; Jenner, 2004b; Draper *et al.*, 2007; Springer *et al.*, 2007; Near 2009). Occasionally, this is attributed uncritically and *a priori* to a spurious morphological tree, with little specific justification (e.g., Hedges & Sibley, 1994; Hedges & Maxson, 1996; D'Erchia *et al.*, 1996). More often, however, authors acknowledge the value of both classes of data, and cede that incongruence may

actually be informative. In all but the most exceptional circumstances (such as in artificial selection experiments or where cladogenesis has occurred in recorded history) one can never be sure that any analytical result matches the 'correct' species tree. Moreover, perfectly bifurcating cladograms may be an oversimplification of actual evolutionary history (Doolittle & Bapteste, 2007; Lopez & Bapteste, 2009; Ragan *et al.*, 2009).

2.2.1 Congruence between Partitions of Morphological Data

While the congruence of signals from different loci is routinely investigated for molecular sequence data, morphological data sets are rarely partitioned explicitly, still less subjected to tests of congruence (although see Smith, 2010; Clarke, 2011; Bennett, 2013 for some recent exceptions). Homogeneity of the morphological signal has often been tacitly assumed (Mayr, 1953; Michener, 1953): the 'hypothesis of nonspecificity' (Sokal & Sneath, 1963). Early attempts to test this assumption applied phenetic methods to data sets of modest dimensions, and yielded equivocal results. Sokal and Sneath (1963), who addressed the issue in the context of vertebrate systematics, stressed the similarity between dendrograms from different sources and concluded that nonspecificity was the rule. Farris (1971), by contrast, emphasised the detailed differences in such cases. If anything, the ascendancy of cladism (*sensu* Hennig, 1966) appears to have further relegated the issue of character congruence for morphological partitions: only a handful of studies have addressed the issue directly (e.g., Sánchez-Villagra & Williams, 1998; Gould, 2001; Song & Bucheli, 2010). With the rise of molecular data, the issue of morphological *versus molecular* incongruence has been much more to the fore (Kluge, 1989), possibly motivated by striking examples of conflict between molecular and morphological cladograms in some groups (Bledsoe & Raikow, 1990; Hillis & Wiens, 2000; Wiens & Hollingsworth, 2000; Pisani *et al.*, 2007; Mayr, 2011a). In the face of these apparently much more invidious difficulties, seeking nuances of signal variation within (potential) partitions of the morphological data has been a low priority. Historically, the focus of research also shifted to taxonomic congruence (*sensu* Mickevich, 1978; Miyamoto & Fitch, 1995) (differences in the implied relationships of taxa on alternative trees), rather than *character* congruence.

When cladograms were predominantly generated manually, the processes of formulating and coding characters were intimately associated with those of tree

construction (Hull, 1990). Practical considerations limited the number of characters that could be analysed, and putative characters were iteratively tested against one another and the coalescing phylogeny (Kitching *et al.* 1998). This almost certainly had the effect of screening out 'noisier' characters, or those contributing greater amounts of homoplasy to the dataset. To the extent that operational definitions of homology are ultimately predicated on the distributions of other characters, it might be considered that this effectively made assessments of probable homology more stringent. With the appearance of faster computer hardware and parsimony programs (Farris *et al.*, 1970), there were less computational restrictions on the size of datasets that could be processed and the complexity of character conflicts that could be resolved. Mooi & Gill (2010) argue that this, coupled with the desire to increase the ratio of characters to taxa, may have encouraged practitioners to incorporate as many characters as possible, with more relaxed 'quality control'. At the same time, there is no imperative for authors to explore the congruence of their characters *a priori* (*sensu* Grant & Kluge, 2003), because computer algorithms will resolve the conflicts, however 'noisy'. Cladistic analyses therefore have two temporally and logically distinct phases (Winther, 2009): character analysis (or the determination of homology: Pinna, 1991), and phylogenetic analysis. Mooi & Gill (2010, p1) contend that current practices are heavily skewed towards the latter, relying too heavily on "algorithms and statistics rather than biology to determine relationships". We do not interpret this as an indictment of the value of methodological advances, but rather as highlighting the comparable dearth of work on evaluating homology prior to analysis.

The concept of 'character congruence' is akin to Sneath & Sokal's (1963) 'nonspecificity', but in an expressly cladistic context. Characters are congruent if their character state trees (*sensu* Estabrook, 1968) are compatible (Le Quesne, 1969), in much the same way that entire cladograms are congruent in the absence of conflicting nodes (Wheeler, 1981).

2.2.2 Previous Quantitative Studies

There is an extensive and mature literature on the use of morphological characters to infer the phylogeny of vertebrate groups, with studies having burgeoned in parallel with the development of modern cladistic methods (Hennig, 1950; Ashlock, 1974). Historically, the focus has been on osteological characters, but morphology also includes non-osteological

‘soft part’ characters such as those from the integument, internal organs, musculature, reproductive organs, gametes, tissue and cellular structures. Many other characters that are not observable morphologically are nonetheless grouped with them in most character lists (ostensibly because they are not conventional nucleotide or amino acid sequence data). These include karyotypes (Faivovich, 2002) , behavioural (Lee & Scanlon, 2002; Faivovich, 2002; Hill, 2005; Li *et al.*, 2007; Spaulding *et al.*, 2009), and ontogenetic sequence (Seiffert, 2007; Simmons *et al.*, 2008) characters. Despite the increasing importance of molecular and genomic data over the last two decades, morphology still makes an invaluable contribution to vertebrate phylogenetics, and is indispensible for the analysis of fossil species. Levels of morphological homoplasy in vertebrate groups are generally lower than those amongst their invertebrate counterparts (Hoyal Cuthill *et al.* 2010), suggesting that the signal quality for vertebrates is relatively high. Moreover, the inferred relationships within many vertebrate clades have altered little with the progression of research time or with the addition of molecular data (albeit with some spectacular exceptions: e.g., Asher *et al.* 2009).

Although cladists occasionally publish variations of trees derived from subsets of their character data (e.g., O'Leary *et al.*, 2003; Poyato-Ariza, 2003; Diogo, 2004 p417-429; Young, 2005; Farke *et al.*, 2011), very few studies have investigated the performance of character partitions quantitatively and systematically using homoplasy indices. The consistency index (CI) is calculated as the ratio of the minimum number of steps a character can exhibit on any cladogram to the minimum number of steps the same character can exhibit on the cladogram in question. The retention index (RI) is a measure of the level of synapomorphy of characters on a given cladogram. A handful of studies have previously compared distributions of consistency (Kluge & Farris, 1969) and retention (Farris, 1989a) indices across characters in two or more partitions. Most recently, Song & Bucheli (2010) inferred these indices for male genital and non-genital characters in insect systematics, finding genital characters to be statistically less homoplasious. Similarly, two other studies compared internal anatomical and external shell character partitions for brachiopods (Leighton & Maples, 2002) and gastropods (Vermeij and Carlson, 2000). The latter paper revealed that shell characters were significantly more homoplastic than internal anatomical characters: a disconcerting finding in a group whose fossils are studied almost exclusively with recourse to hard part characters. In a study of hedgehogs, Gould (2001) reported significantly higher consistency indices for dental characters compared with the remainder of his dataset. However, he also found that the optimal trees inferred

from the dental characters alone were seriously at odds with those implied by the other characters, and he cautioned against the uncritical use of the former. Most ambitiously, Sanchez-Villagra and Williams (1998) compared the consistency indices between dental, cranial and postcranial character partitions of eight mammalian data sets. These were of modest proportions (an average of eleven taxa), but they reported no significant differences for any of their comparisons. Unfortunately, this general approach is complicated by differences in the sizes of the dataset partitions, coupled with the inverse correlation between both consistency and retention indices and the number of taxa (Archie, 1989; Sanderson & Donoghue, 1989; Klassen *et al.*, 1991; this paper). We therefore adopt a variety of allied methods in this study.

2.2.3 Why Examine the Congruence of Cranial and Postcranial Partitions?

Character analysis is often the least well explored aspect of systematic analyses (Pogue & Mickevich, 1990). Amongst vertebrate systematists, it is often asserted that cranial and postcranial characters convey signals of differing quality (Ward, 1997; Collard *et al.*, 2001; Naylor & Adams, 2001; Finarelli & Clyde, 2004). However, the evidence for this is piecemeal and largely anecdotal, with few attempts to quantify putative differences. Many practitioners take a more or less even-handed approach to sampling characters (Sánchez-Villagra & Williams, 1998), attempting to avoid describing a disproportionate number from any one anatomical region (Sokal & Sneath, 1963). However, even where potential characters are reasonably homogeneously distributed throughout the body, “certain body regions and organs still hold a considerable mystique for taxonomists as classificatory tools, while others are neglected” (Sokal & Sneath, 1963; page 85). For example, Arratia (2009) notes that actinopterygian systematists focus their attention significantly toward cranial characters, despite rich seams of underexploited data within the fin rays and fulcra. In this study, we apply a variety of methods to explore differences in the strength and character of phylogenetic signals in cranial and postcranial partitions of 62 published vertebrate data sets.

We address the following questions: 1. Is there a significant difference between the number of cranial and postcranial characters in our sampled data sets? Received wisdom

holds that the cranium is a richer source of phylogenetic data that the postcranium (Sanchez-Villagra & Williams, 1998). 2. Are levels of homoplasy in cranial character partitions lower than in postcranial character partitions (Sanchez-Villagra & Williams, 1998), and are any observed differences more than simply a function of differing numbers of characters within these partitions? 3. Is there more conflict between cranial and postcranial characters than we would expect, and are the relationships inferred from them less similar than we might predict?

2.3 Materials and Methods

Phylogenetic data sets published between 2000 ad 2010 were sourced from the peer-reviewed literature. We restricted our focus to discrete character morphological matrices composed entirely of vertebrate taxa, and analysed using equal weights maximum parsimony. Matrices were initially garnered from Brian O'Meara's *TreeBASE* mirror (O'Meara, 2009), Graeme Lloyd's online collection of non-avian dinosaur matrices (Lloyd, 2009), *MorphoBank,* (O'Leary & Kaufman, 2011) directly from authors, or from the original papers. We then filtered these by removing matrices with no cranial or postcranial characters, in addition to those with fewer then eight taxa or partitions with fewer than eight parsimony-informative characters (for reasons of statistical power). We interpret cranial characters here as those pertaining to the skull (cranium plus mandible and dentition) rather than as *just* those of the cranium. In cases of taxonomic overlap (or where one matrix was derived from an earlier one), we retained only the most inclusive (usually the most recent) dataset if the fraction of terminals in common was 50% or more of the size of the smaller matrix. A modest number of matrices were also excluded because they contained more than 200 taxa or more than 1,000 characters (e.g., for Squamata; Conrad, 2008 and derivative papers). These were highly atypical, and could not have been analysed with comparable rigour in a tractable time with our methods.

Our resulting sample comprised 62 matrices, spanning all major vertebrate groups, and sampled at a variety of taxonomic levels. A minority of these data sets contained a small number of characters that were not strictly morphological (e.g., character 618 relating to habitat choice in the matrix of Spaulding *et al.*, 2009). These were removed prior to any further analysis. We also removed phylogenetically uninformative taxa within partitions using the principles of safe taxonomic reduction implemented in *TAXEQ3* (Wilkinson, 1995b). Additionally, parsimony-uninformative characters were removed

because they have undesirable effects on many of the tests that we subsequently performed (Carpenter, 1988; Sanderson & Donoghue, 1989; Bryant, 1995; Lee, 2001b). The number and percentage of cranial and postcranial characters in each matrix were recorded. Wilcoxon signed-rank tests were used to assess the significance of differences between these means in different groups.

2.3.1 Do Cranial and Postcranial Characters Imply Different Levels of Homoplasy?

Consistency and retention index comparisons. – All phylogenetic analyses were performed using PAUP* 4.0b10 (Swofford, 2002), using equally weighted parsimony analysis. We also reproduced any preferred assumptions concerning character order, as well as assumptions concerning character polarity and rooting. Empirically, we determined that 20 random addition sequence replicates followed by TBR branch swapping holding up to 1000 trees at each cycle (and limited to retaining a maximum of 10,000 trees overall) were effective at recovering the set of MPTs reported by the original authors in each case. We therefore used these settings for all subsequent analyses. Such heuristic searches were a necessary compromise given constraints of computational time and memory. Cranial and postcranial character partitions were specified using charset commands. There are two obvious ways in which to calculate differences in mean/median consistency indices (ci; Kluge & Farris, 1969) and retention indices (ri; Farris, 1989) for characters in partitions of a dataset, but there are drawbacks to both. The first (and usual) approach is to find the optimal tree or trees for all characters simultaneously (the global MPT(s) on the principle of total evidence; Kluge, 1989) and to take mean values for characters reconstructed on this/these (Sánchez-Villagra & Williams, 1998; Song & Bucheli, 2010). Where characters in the two partitions contain different levels of homoplasy but support the same tree(s), this approach is relatively straightforward. However, the situation is more complex when the partitions support different trees, and especially when these partitions are also of different sizes. In this case, with *no* differences in the levels of character conflict *within* partitions (and all other things being equal), the larger partition is more likely to determine the overall pattern of relationships.

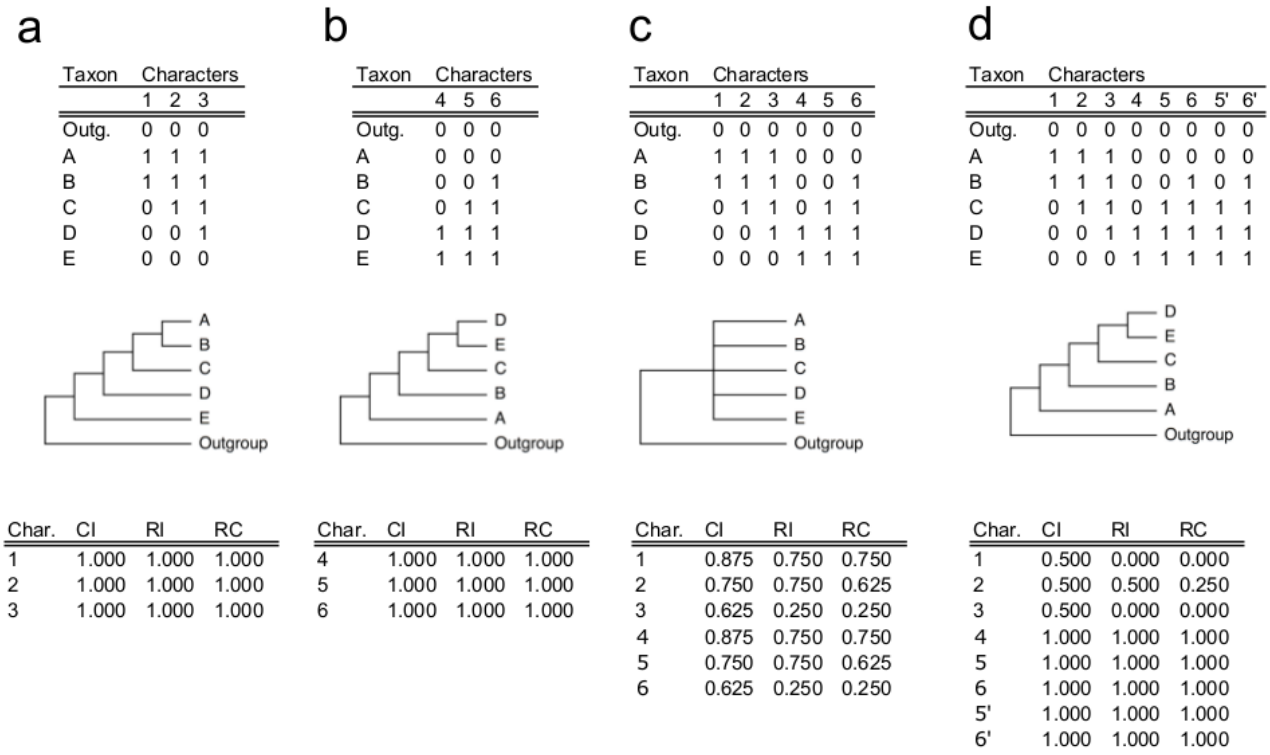


Figure 2.1 a & b) A theoretical example. Characters sampled from different anatomical regions can yield radically different most parsimonious trees (MPT) when analysed in isolation. In both cases, there is no homoplasy within either region (characters 1-3 or characters 4-6), and a single MPT results in each case. c) Combining the data from both partitions (characters 1-6) yields four MPTs, the strict consensus of which (illustrated) is entirely unresolved. Character statistics have been averaged over the four trees. d) Two additional characters (5' and 6') are sampled from the same region as 'b', and these have the same distribution as 5 and 6 respectively. Analysis of all characters now reveals a single MPT with relationships identical to those in 'b' (characters 4-6). Characters 4-6 , 5' and 6' contain no homoplasy: all conflicts are resolved with a cost to characters 1-3. In this case, the MPT is identical to the result that would be obtained by a clique analysis (sensu Le Quesne 1969).

This also means that the characters in the larger partition are likely to have higher ci values on average (Fig. 2.1). Notwithstanding, evaluating and resolving differences in numbers of characters supporting incompatible hypotheses is an integral part of total evidence analysis (and of parsimony in general). However, another part of the rationale is to combine all available data, or at least to sample in an unbiased manner from the universe of possible characters. Because systematists tend to concentrate on particular body regions, there is the potential for self-reinforcement: the undersampled characters will actually have lower ci values upon inspection (because other body regions dominate the overall phylogenetic signal), and therefore may be less likely to receive attention in future analyses. Mean partition ci calculated in this way therefore has the potential to

reflect sampling intensity. The second approach to measuring homoplasy is to investigate the performance of the characters within each partition analysed independently. While this will reveal information about homoplasy within partitions (rather than in one partition relative to the entire sample), the ensemble CI and ensemble RI (and therefore ci and ri for individual characters) are known to be influenced by data set dimensions (Archie & Felsenstein, 1993; Archie, 1996). This effect is particularly pronounced for the number of taxa, where the correlation is strong and negative. This is not a problem for comparisons within data sets (as here), because the number of taxa is always constant. However, the number of characters also has a much less marked but negative impact: one character cannot conflict with itself, while two are less likely to conflict than twenty (Fig. 2.2). This bias (for CI) operates in the opposite direction to that observed above (for mean ci in a global optimization).

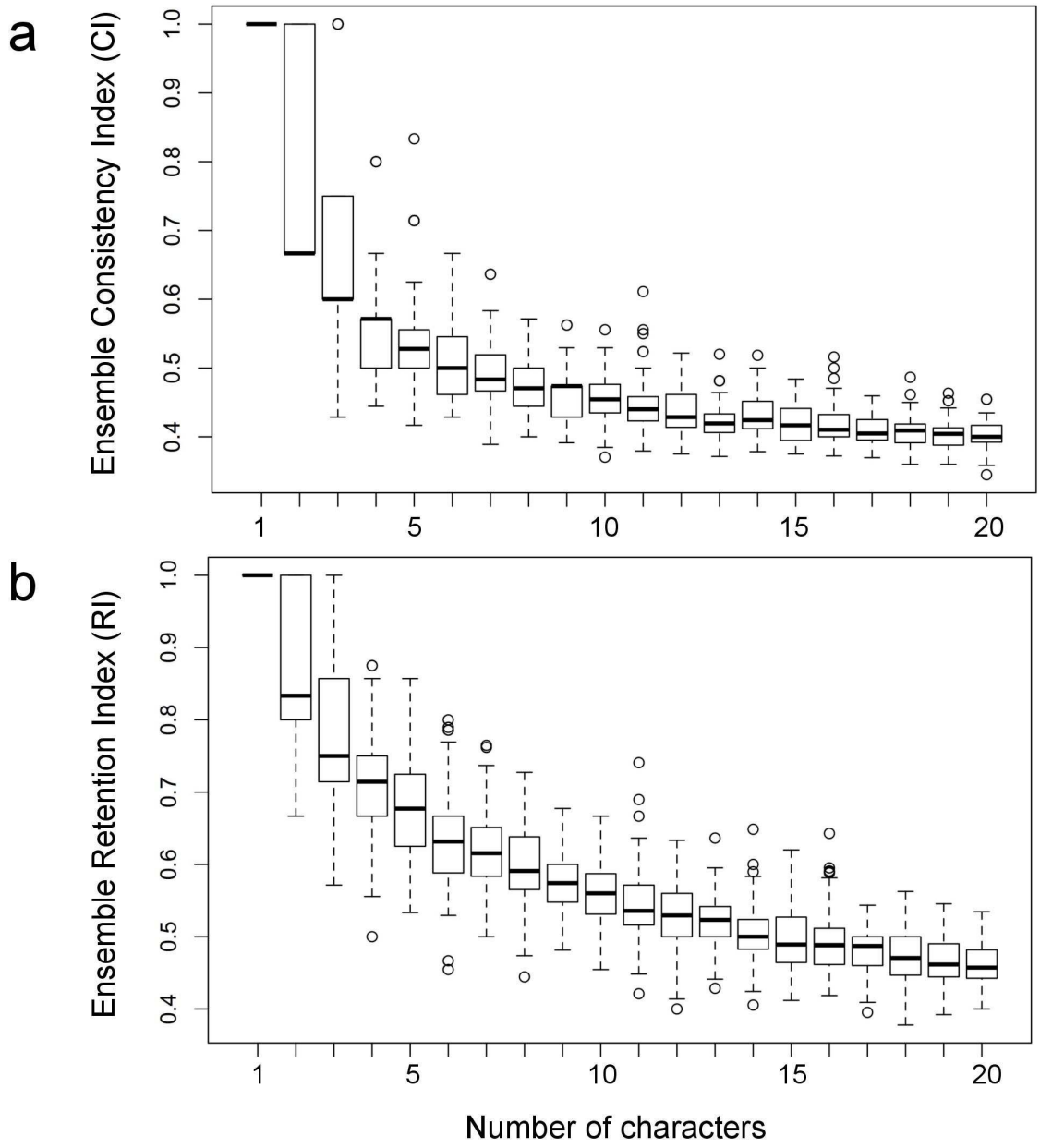


Figure 2.2 The number of characters in a data matrix influences probable indices of homoplasy. a) The ensemble Consistency Index (CI). b) The ensemble Retention Index (RI). Simulations for the trivial case of 10 taxa, with between 1 and 20 characters. All states were randomly assigned 0 or 1 with equal probability. Box and whisker plots summarise 100 replications for each number of characters.

There are three ways in which differences between these indices can be tested. For individual data matrices, Mann-Whitney or t-tests can be applied to character ci and ri values, with the null that these have a similar median or mean in the two partitions. For the more general comparison across all 62 matrices simultaneously, Wilcoxon signed ranks or paired t-tests can be used to test the nulls that (either) the median/mean ci or ri in cranial and postcranial partitions were similar, or that the median/mean CI and RI indices for the two partitions were similar. As discussed above, however, we note that all of these outcomes are differently and undesirably influenced by partition size.

We also note the potential for a test based on the distributions of values obtained from bootstrapped samples, controlling for differences in character number by repeatedly subsampling both/all partitions at the size of the smallest. We have not implemented such a test here.

Homoplasy Excess Ratio (HER) indices. – The homoplasy excess ratio (HER; Archie & Felsenstein, 1993) was proposed as an adjunct to the ensemble consistency index (CI), and argued to be relatively immune to its worst shortcomings. Central to the calculation of the index is a randomisation procedure that operates by repeatedly permuting the assignment of character states within characters but across taxa, thereby disrupting any phylogenetic signal. A large number of randomised matrices are then analysed under maximum parsimony in order to obtain a distribution of tree lengths. This is similar to the procedure implemented by the permutation tail probability test (PTP; Faith & Cranston, 1991). The limitations of the latter test have been rehearsed at length elsewhere (Wills 1999). However, the HER differs fundamentally from the PTP, because it does not use this distribution as the means to test a null. Rather, the mean value for randomised matrices (MEANNS) is used as an estimate of the expected tree length for matrices of the same dimensions and with identical frequency distributions of states as the original. The HER is then calculated as:

$$HER = (MEANNS - L) / (MEANNS - MINL)$$

where L is the optimal length of the original dataset and, MINL is the minimum possible length of the dataset. MINL was calculated as the total number of character states in the entire matrix, minus the number of characters. MEANNS was estimated in *PAUP** using the *permute* command with 999 replicates and the search parameters used on the original dataset.

The HER is calculated as a single value for a given block of data: there is no formally-proposed analogue of the index for individual characters (as the ci is a an analogue of the CI). We note that such a test would be possible by permuting the state assignments for individual characters, but we do not explore this here. The HER has therefore been calculated for the cranial and postcranial halves of each matrix when analysed in isolation. The properties of the HER mean that differences in the sizes of partitions are largely controlled for. We then tested for differences in partitions across all 62 data sets using the Wilcoxon signed ranks test.

2.3.2 Is there More Conflict Between Cranial and Postcranial Characters than we Might Expect?

Incongruence Length Difference (ILD) test. – To assess the significance of congruence between whole character partitions as measured by optimal tree length, the ILD test (Mickevich & Farris, 1981; Farris *et al.,* 1995a; Farris *et al.,* 1995b; Barker and Lutzoni, 2002) was applied to the matrices in *PAUP** using the *hompart* command, with 999 replicates (Allard *et al.,* 1999a,b). Heuristic search settings were specified using tree bisection-reconnection (TBR), 10 random addition sequence replicates, holding up to 1000 trees at each cycle, limited to holding a maximum of 10,000 trees overall. These tests were run on a high performance computing cluster (Bioportal; Kumar *et al.,* 2009).

The ILD score is given by $L_{AB} - (L_A + L_B) / L_{AB}$ where L_{AB} is the optimal tree length (in steps) of the simultaneous analysis of both partitions together (the total evidence analysis). L_A is the optimal tree length of an analysis of just partition A, and likewise L_B is the optimal tree length of an analysis of just partition B (Fig. 3). To determine the significance of the observed ILD score, random partitions of the same size (number of characters) as the specified partitions are also generated to yield a distribution of randomized ILD scores. Thus, the ILD test is a randomization test (*sensu* Kempthorne, 1952), that compares the significance of a particular score relative to the scores of a set of randomly permuted replicates of the same dataset. Given the nature of phylogenetic data, the suitability of this test has been questioned on a variety of grounds (Dolphin *et al.* 2000; Hipp *et al.* 2004; Ramirez 2006; reviewed in Planet 2006). Despite this, the ILD test remains commonly used to compare the congruence of data partitions). We did not apply the arcsine

transformation suggested by Quicke *et al.* (2007) because they justified their correction on the basis of empirical and simulated *molecular* data, whilst here we use morphological data which has different statistical properties. Morphological data matrices are mostly composed of binary or three-state characters and contain ordered characters, whilst molecular data is typically of four-state characters (e.g. GTCA) and these are never ordered in linear sequence between states. To investigate the robustness of the ILD test to taxonomic sampling, we also performed first-order taxon jack-knife ILD tests (cf. Planet & Sarkar, 2005, but not using their scripts) on a selection of the smaller taxon data sets. This enabled global partition incongruence (all or most taxa) to be distinguished from local incongruence (caused by individually highly incongruent taxa).

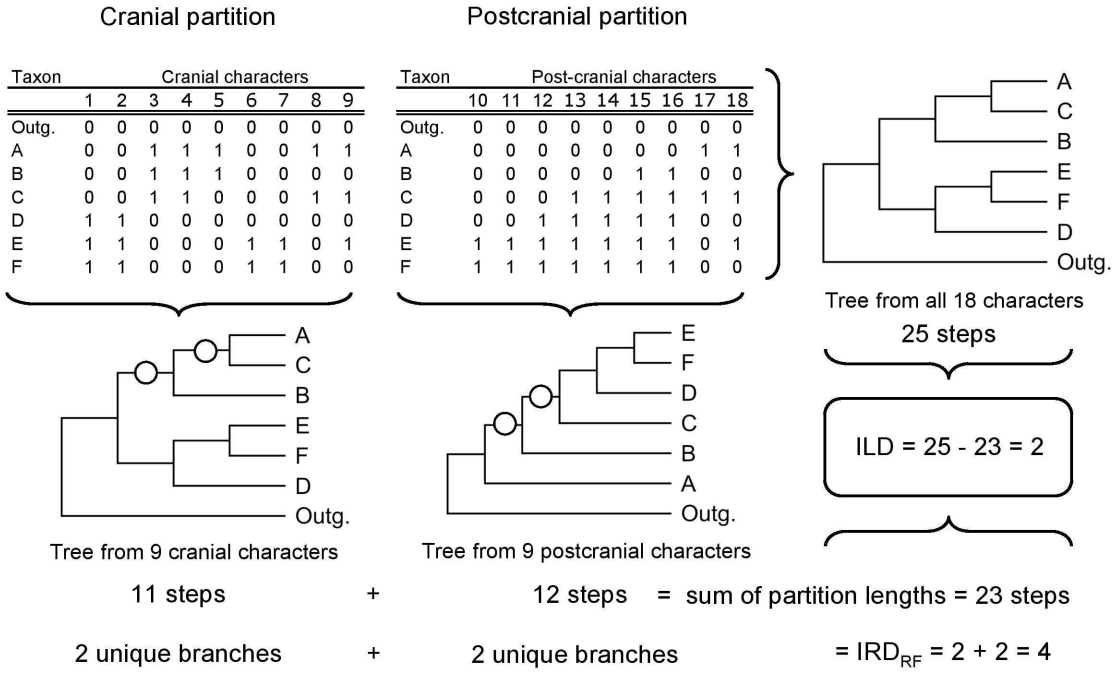


Figure 2.3 Calculation of two inhomogeneity metrics for 'cranial' and 'postcranial' partitions of a hypothetical data set. In this example there are equal numbers of cranial (1-9) and postcranial (10-18) characters, but this need not be the case. For the Incongruence Length Difference (ILD) measure, maximally parsimonious trees (MPTs) are inferred from the cranial and postcranial partitions of the data independently. The summed lengths of these trees (11 steps + 12 steps) is the sum of partition lengths (23 steps). In parallel with this, an MPT is inferred from both partitions analysed simultaneously. This tree is longer (25 steps) than the sum of partition lengths (23 steps), and the difference between them is the ILD (25 – 23 = 2). The ILD represents the reduction in homoplasy afforded by the isolation of the two partitions (two extra steps are needed when the partitions are combined). For the Incongruence Relationship Difference (IRD) measure, the

branching structure of the cranial and postcranial partition MPTs are compared (rather than their lengths) using one of several possible tree-to-tree distance metrics. Here, we illustrate the symmetric difference distance (RF) of Robinson and Foulds (1981) (so the metric is the IRDRF). Open circles mark branches in either the cranial or postcranial MPT that are absent from the other. The tally of these unique branches on both trees is the RF (2 + 2 = 4). Some background level of ILD or IRD is anticipated wherever a data set contains homoplasy. In order to interpret these observed metrics, therefore, we need to know what values would be expected for partitions of similar data sets in similar proportions. Random character partitions are used to generate null distributions for both the ILD and IRD, and observed valued deemed significantly different from the null if they lie in some specified fraction of the tails.

Do Cranial and Postcranial Characters Support Different Trees?

Templeton, winning sites and Kishino-Hasegawa tests. – All of these tests can be used to assess whether a given matrix offers significantly more support for one tree compared with another. In order to interpret the statistics that they generate straightforwardly, the alternative trees should be specified *a priori* rather than from an analysis of the matrix. In many applications, however, it is common to compare an optimal tree with an alternative to determine whether the latter is significantly worse. In this context, the trees inferred from the cranial and postcranial partitions were the suboptimal alternatives to MPTs from the postcranial and cranial partitions respectively. We also tested trees from individual partitions against the entire dataset. We note that these applications of the tests may be problematic (with a high type I error rate) because we are comparing an optimal with a suboptimal tree by definition (rather than two alternative trees derived independently from our data) (Goldman *et al.*, 2000).

The Templeton test operates by calculating the length of each character on both the optimal and the alternative tree (Templeton, 1983). These paired values are then subjected to a one-tailed Wilcoxon test (Siegel & Castellan, 1988). A one-tailed test was used because although the steps contributed by an individual character can be fewer on the suboptimal than the optimal tree, the contributions summed over all characters can only be greater. The winning-sites test (Prager & Wilson, 1988) is very similar to the Templeton test, except that it ignores the magnitude of differences and uses only counts of 'winning-sites' (parsimony-informative characters that fit more parsimoniously on one tree-topology than the other). These two sums were then analysed with a one-tailed binomial test.

Finally, the Kishino-Hasegawa (KH) test (Kishino & Hasegawa, 1989; more specifically 'test priNPncs' according to Goldman *et al.* 2000, p658) also computes site-wise differences between trees, but makes the additional assumption that differences between sites are normally distributed. These site-wise differences of fit were then tested with a one-tailed, paired t-test: as such the KH test is effectively a parametric analogue of the Templeton test.

The KH test has a related family of variants (Swofford *et al.,* 1996; Shimodaira & Hasegawa, 1999; Buckley *et al.,* 2001; Shimodaira, 2002) commonly used to compare trees generated in a Maximum Likelihood framework. The Shimodaira-Hasegawa test is designed for the particular case where an optimal tree is compared with a sub-optimal alternative. Although this can be applied in a parsimony context (Near *et al.,* 2003) we are not aware of a straightforward implementation.

In practice, many partitioned and entire data sets yield more than one MPT. Where numbers are small, it would be feasible to test all alternatives. However, in many cases the numbers preclude this. We have therefore used majority rule consensus trees (including only compatible groupings with greater than 50% support). We are aware of the limitations of this approach; specifically that majority rule consensus trees need not lie at the centre of the 'tree spaces' defined by their fundamentals (see Fig. 2.3). We also note that when applied to comparisons of trees from partitions versus those from entire data sets (and *all other things* being equal) the 'entire' tree is more likely to be similar to that from the partition with the greater number of characters. All tests were implemented in *PAUP** (Swofford, 2002).

Topological Incongruence Length Difference (TILD). – The TILD test (Wheeler, 1999) operates in a manner analogous to the ILD test, but is applied to a matrix representation of the branching structure of the optimal trees from the data partitions (rather than to the partitioned character data itself). Cranial and postcranial partitions are analysed independently, and a majority-rule (>50%) consensus tree is generated for each of them. Each consensus is then translated into a matrix of group inclusion characters (GIC; Farris, 1973; also known as MRP coding, Baum, 1992; Ragan, 1992) that convey the same information as the cladogram branching structure. The GIC matrices for the partitioned analyses A and B are re-combined and subjected to a conventional ILD test.

The Incongruence Relationship Difference test (IRD); a new test of the congruence of relationships (Wills pers. comm)

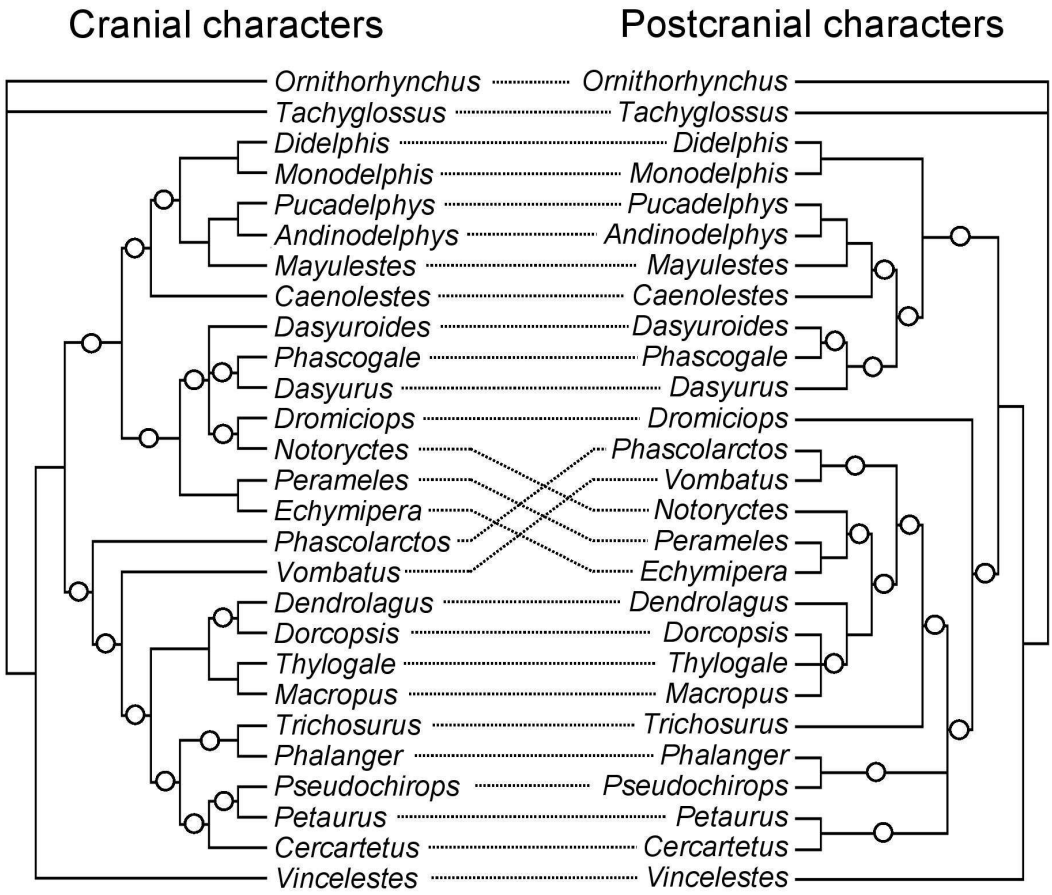


Figure 2.4 Most parsimonious trees derived from the 103 cranial and 139 postcranial characters in the mammalian data of Beck (2008). Four cranial and eight postcranial trees are summarized here as majority rule consensuses. Nodes labeled with circles are unique to one or other tree (those unlabelled are common to both). The Robinson Folds (RF) distance between the two is simply the sum of unique nodes (15+15=30). The tanglegram was computed using Dendroscope (Huson and Scornavacca, 2012). In this case, the incongruence length difference (ILD) test for partition homogeneity returned a highly significant result ($p=0.004$) whereas all tested variants of our incongruence relationship difference (IRD) test were not significant ($p \geq 0.11$). See text for further explanation.

Much like the ILD test, this is a randomisation-based test (Fig. 2.3). However, partitions are compared via the distances between the optimal trees that result from them, rather than via tree length (ILD) or a matrix-representation of topology (TILD). There are many possible tree-to-tree distance measures including symmetric difference (RF; Bourque, 1978; Robinson & Foulds, 1981) quartets distance (QD; Estabrook *et al.,* 1985), nearest neighbour interchange distance (NNID; Waterman & Smith, 1978), nodal distance (Bluis & Shin, 2003), maximum agreement subtree distance (Goddard *et al.*, 1994; de Vienne *et al.*, 2007), transposition distance (Rossello & Valiente, 2006) subtree prune and regraft distance (SPR; Goloboff, 2008), and path-length difference (PLD; Zaretskii, 1965; Williams & Clifford, 1971). For reasons of familiarity (they are among the most well characterised; e.g. Steel & Penny, 1993) and ease of use (they are already implemented in *PAUP**) we chose to use both the symmetric difference (RF) (Fig. 2.4) and the agreement subtree metric (AgD1) as our measures of tree-to-tree distance. We note that all other implementations are possible

Briefly, a heuristic search on each partition was specified using tree bisection-reconnection (TBR), holding up to 500 trees at each cycle, and limited to holding a maximum of 10,000 trees overall. All MPTs from the analysis of each partition were saved and then compared to each other in two different ways. (1) 'Nearest neighbours' (IRD_{NND}) for up to 1,000 trees in each partition: the mean of the minimum distance between each tree in one set, compared with the trees in the other (and *vice versa*) (Cobbett *et al.*, 2007). (2) The distance between the 50% majority-rule consensus trees (from up to 10,000 fundamentals) for each partition (IRD_{MR}) (Fig. 2.5). We then generated random partitions of the original data in the original proportions, and repeated the above exercises in order to yield a distribution of randomized partition tree-to-tree distances. Distances for the original partitions were deemed significantly different from this distribution when they lay in its 5% tail. We implemented nearest neighbour and majority rule variants using symmetrical difference distances (IRD_{NND+RF} and IRD_{MR+RF} respectively) and the majority rule variant with maximum agreement subtree distance ($IRD_{MR+AgD1}$); three statistics in total. Most of our p-values were derived from just 99 replicates (in contrast to the 999 used for ILD and TILD) because of time and computational constraints.

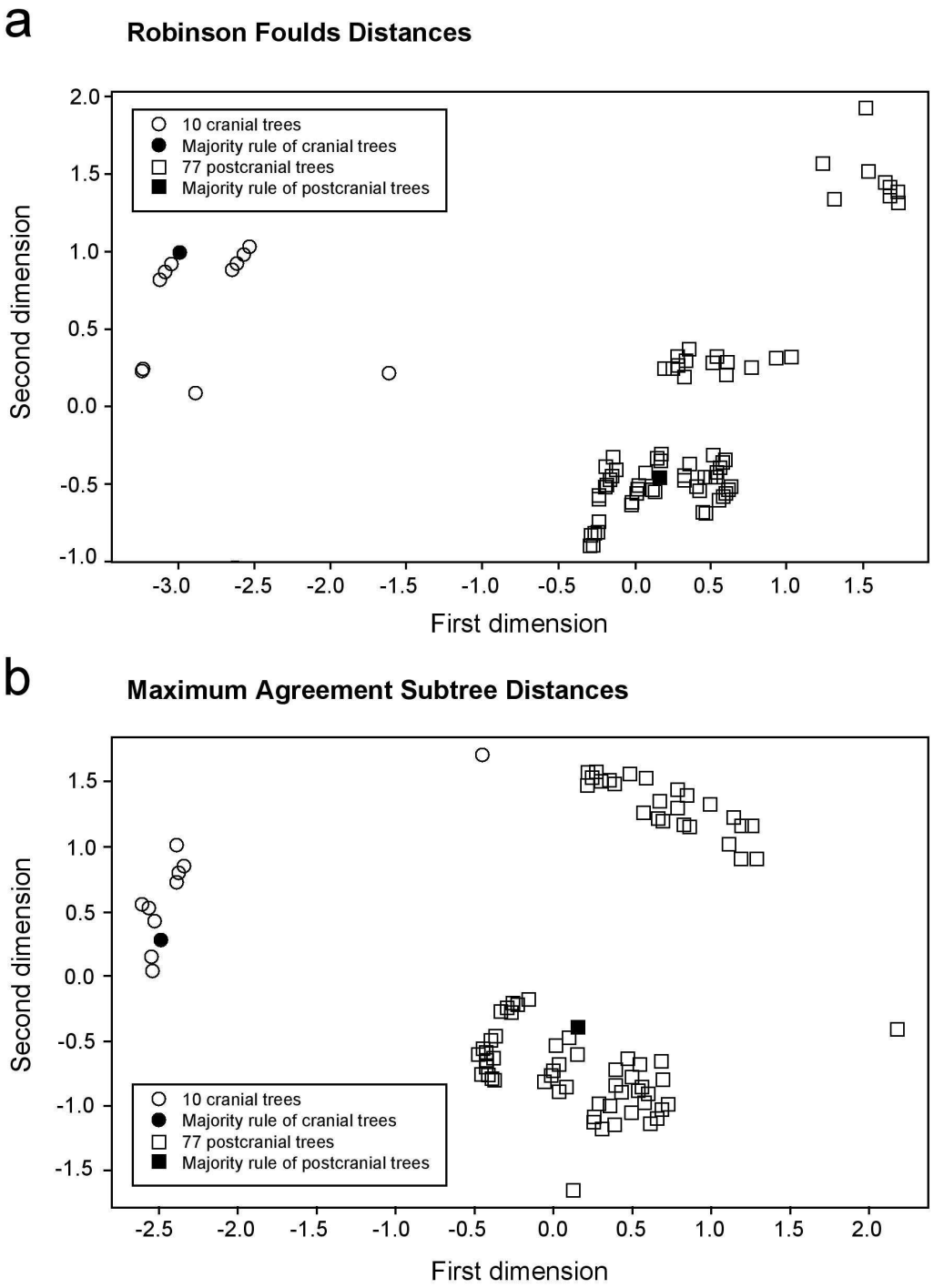


Figure 2.5 Tree-to-tree distances for cranial and postcranial partitions of the mammalian data of Pujos (2007). Distance matrices have been plotted in two dimensions using non-metric multidimensional scaling (NMDS), and rotated using principal components analysis (PCA). Circles indicate cranial trees and squares indicate postcranial trees. Open symbols denote original MPTs, filled symbols (black) denote majority rule trees. a) Robinson Foulds (RF) distances. b) Maximum agreement subtree distances (AgD1).

Tests not performed in this study. – Rodrigo *et al.* (1993) proposed three interrelated tests to investigate differences in relationships directly. The first of these determines whether the symmetrical difference distance (RF: Robinson & Foulds, 1981) between sets of MPTs from independent analyses of the two data set partitions is distinguishable from the distribution of RF distances between a large sample of pairs of random trees. At face value, this is an extremely easy test to pass. It requires only weak congruence between the two data set partitions, and does not assess directly whether the partitions behave as though they are randomly sampled from the same universe of characters. Rodrigo *et al's* second test *does* compare the partitions directly (rather than with reference to random trees), and determines if there is any overlap between the MPTs derived from the two partitions upon perturbation of those partitions. Specifically, both partitions are subject to a bootstrapping analysis, and the MPTs obtained are compared. The null hypothesis is that the two data partitions result from the same phylogenetic history, and that there should therefore be "significant" overlap between the two sets (although this is not defined). The protocol is unclear, but suggests that if there are no trees common to both sets then the null can be rejected. Conversely, even one tree common to both sets leads to acceptance of the null. As noted by Lutzoni (1997), this is problematic because the probability of encountering common trees changes with the bootstrap parameters, especially the number of replicates (Page, 1996). The third test is allied to the second, but is more robust. If both partitions of the data set are sampled randomly from the same universe of characters (the null expectation), then the symmetric difference distances between pairs of MPTs from the two partitions should be similar to the distances between pairs of trees obtained by bootstrapping individual partitions. Although a useful test, it may have limitations, particularly where the partitions of the data set are of very different sizes, and especially where the number of characters in the smaller partition is also small relative to the number of terminals. In such cases, bootstraps of the smaller partition may consistently yield poor resolution and low RF distances between trees *within* this partition (Page, 1996). The IRD test proposed above controls this partition size difference.

2.4 Results

Table 2.1 Meta-summary table classified into significant and non-significant results, for comparisons of cladistic data matrices partitioned into cranial and postcranial. See Appendix 2.1 and online materials for fuller supplementary data.

	No. Datasets Significantly Different Between Partitions (<0.05)	**Exemplars**
ILD	18 (29%)	Choristodera (Matsumoto *et al*, 2009) ILD = 0.002
TILD	50 (81%)	Amniota (Hill, 2005) TILD = 0.001
IRD_{NND+RF}	20 (32%)	Ankylosauria (Osi *et al*, 2009) IRD_{NND+RF} = 0.010
IRD_{MR+RF}	18 (29%)	Plateosauria (Smith & Pol, 2007) IRD_{MR+RF} = 0.020
	Significance of difference across all 62 datasets, between partitions	**Inference**
CI	Significant (paired t = -2.2278, p = 0.0296)	Postcranial significantly higher CI than cranial
RI	Not Significant (paired t = 1.0489, p = 0.2984)	Cranial higher RI on average, but not significant
Mean ci	Not Significant (paired t = -1.3422, p = 0.185)	Postcranial per character ci higher but not significant
Mean ri	Significant (paired t = 4.7324, p =1.36e-05)	Cranial per character ri significantly higher
HER	Significant (paired t = 2.042, p = 0.0455)	Cranial HER significantly higher than postcranial HER

2.4.1 Homoplasy and Resolution in Cranial and Postcranial Data Partitions

Across our sample of 62 data sets, cranial partitions had significantly more characters (median = 66) than postcranial partitions (median = 54.5) (Wilcoxon signed ranks; V = 1398, p = 0.0032) (Appendix 2.1). This difference would have been more marked were it not for the inclusion of multiple ornithological studies (e.g., Bourdon *et al.,* 2009; Ksepka, 2009; Worthy, 2009) that are counter biased towards greater numbers of postcranial characters. We also selected matrices containing *relatively* balanced numbers of characters *a priori*, making a significant difference less likely.

Differences in numbers of characters across partitions mean that it is not straightforward to compare ensemble consistency indices (CI) and retention indices (RI). Nonetheless, we observe that the mean CIs for cranial and postcranial characters across all 62 data sets were significantly different (paired t = -2.2278, p = 0.0296), with postcranial partitions ($\bar{x}$ = 0.549) having slightly higher self-consistency than cranial partitions ($\bar{x}$ = 0.512) overall. This difference may, in part, be a function of the smaller average number of characters within postcranial partitions. Indeed, using the mean partition (per character) ci index across all matrices revealed a difference in the same direction ($\bar{x}$ = 0.560 and 0.540 for postcranial and cranial partitions respectively), but this was not significant (paired t = -1.3422, p = 0.185). Mann-Whitney tests of cranial and postcranial ci values *within* our 62 data sets yielded fourteen significant ($p < 0.05$) results (three or four might be expected). Ten of these fourteen had higher means (less homoplasy) for postcranial partitions; an insignificant bias (binomial test, p = 0.180) that was nonetheless in general agreement with the other ways of expressing differences in CI and ci.

No significant difference was found in the test of cranial versus postcranial ensemble retention indices (RI) across data sets (t = 1.0489, p-value = 0.2984) but per character retention indices (ri) showed a highly significant difference (t = 4.7324, p = 1.36e-05) in the *opposite* direction (higher values in cranial partitions) from the CI and ci. Higher Homoplasy Excess Ratio (HER) values were also found in the cranium; the mean value for cranial partitions (0.5675) was significantly higher than that for postcranial partitions (0.5207) (t = 2.042, p = 0.0455). Our results are equivocal, therefore, depending upon precisely how homoplasy is measured.

The strict consensus fork index (CFI; Colless 1980; Table 2.1) also demonstrated that the number of nodes resolved in cranial partitions is greater than that in postcranial

partitions. Only 16 cranial partitions resulted in a strict consensus with no resolution (CFI = 0), compared with 25 postcranial partitions. All partitions for all data sets had a CFI greater than zero for their majority-rule (>50%) consensuses. A comparison of majority-rule CFIs between cranial and postcranial partitions reveals that the former give significantly better resolution (paired Wilcoxon; V = 1190.5, p=0.004).

2.4.2 Congruence between Cranial and Postcranial Signals (ILD tests)

When originally described, the ILD test was used with a standard significance level of 5% (0.05). At this level, 19 of our 62 data sets had significant character incongruence between cranial and postcranial partitions (Table 2.1). Whilst this significance level remains the most widely quoted, some have suggested more stringent significance levels should be used (e.g., Cunningham, 1997a). Applying stricter significance levels, we reject the null for just 16 ($p < 0.010$) and 8 ($p < 0.001$) of our data sets.

The Similarity of Relationships Implied by Cranial and Postcranial Partitions

TILD and IRD tests. – 50 of our data sets (>80%) had partitions for which majority-rule trees yielded significant TILD tests results ($p \leq 0.05$); 39 of these also had $p = 0.001$. We used 100 replications for our IRD tests because of time constraints. Using the nearest neighbour procedure with symmetrical difference distances (IRD_{NND+RF}), 20 of our data sets (32%) had significantly incongruent relationships implied by cranial and postcranial partitions ($p \leq 0.05$) (Table 1). Using majority rule trees and symmetrical difference distances (IRD_{MR+RF}) yielded highly similar but not identical results ($r = 0.80$, $p = 6.21 \times 10^{-15}$); 18 data sets (29%) had significantly incongruent relationships. There are two reasons for the differences. Firstly, majority rule trees embody most frequent relationships rather than 'average' relationships (Fig. 3), so the two tests address different questions. Secondly, the IRD_{NND+SR} was calculated using up to 1,000 trees from each partition (a maximum of 499,000 tree-to-tree distances for each replicate), but these may not offer representative samples of all islands of MPTs where the latter are very numerous.

The analogous tests using maximum agreement subtree distances (AgD1) were only implemented for majority rule trees, since these distance computations were very much slower in *PAUP**, encountered segmentation faults (in beta 10) and became

prohibitively time consuming for thousands of neighbour comparisons. The $IRD_{MR+AgD1}$ test revealed significantly different ($p \leq 0.05$) majority rule trees for seventeen data sets. Correlation with p-values from the IRD_{MR+RF} test were significant but not especially high (Kendall's $\tau = 0.1895$, $p = 0.0350$) because the two metrics address different aspects of tree-to-tree distance.

Partitioning data sets into four broad taxonomic groups (Mammalia, Avemetatarsales/Ornithodira, fishes and 'other tetrapods') revealed some striking differences, albeit with modest sample sizes. In particular, fishes were more likely to have congruent cranial and postcranial partitions than the tetrapod groups.

Templeton, winning sites and KH tests. – These tests are not straightforward to interpret, because rather than comparing two alternative trees with an independent data set, we were here comparing the optimal tree from a given data set with an alternative (suboptimal by definition). This will yield a high rate of type I errors (Planet, 2006). Moreover, these comparisons were necessarily mediated via consensus trees, which may not be included in the set of fundamentals and may therefore also be suboptimal. Unsurprisingly, therefore, the majority of the tests reported significant differences (see online Appendix 1). Results using majority rule and strict consensus trees were very similar. We summarize these here as the tally of data sets for which optimization of both partitions *and* the entire data set onto the cranial and postcranial trees yielded significant p-values (i.e., where all three comparisons yielded $p \leq 0.05$). For both Templeton and KH tests (both using the magnitude of step differences per character), 38 data sets (61%) had a maximum $p \leq 0.05$ using majority rule trees, while 41 (66%) had a maximum $p \leq 0.05$ using strict consensus trees. The winning sites test (utilizing only the direction of the step differences) was a little more conservative, reporting a maximum $p \leq 0.05$ for 31 data sets using the majority rule and 38 using strict consensus trees.

As above, these statistics consistently reported more congruent cranial and postcranial partitions for fishes than for the terrestrial groups. However, these differences were not significant with the possible exception of the winning sites test with majority rule trees, which was marginal ($G = 6.8299$, χ^2-squared df = 3, p-value = 0.0775).

2.5 Discussion

Cranial and Postcranial Partitions Contain Similar Levels of Homoplasy

We are cautious when interpreting consistency indices in data set partitions, because these partitions seldom comprise identical numbers of characters. Moreover, there is a significant bias toward higher numbers of cranial characters across our sample of data sets. On one hand, the ensemble CI for a partition optimized parsimoniously in isolation is 'biased' by the number of characters (Fig. 2.2); larger partitions can be shown to have a lower CI in random simulations (cf. Sanchez-Villagra & Williams, 1998; Song & Bucheli, 2010). We did, indeed, find a significant difference in ensemble CI between our partitions (paired t = -2.2278, p = 0.0296), with postcranial partitions having higher values overall. On the other hand, if both partitions are analysed simultaneously, then mean or median per character ci values are likely to be higher in the larger partition; at least in the hypothetical case where the partitions contain conflicting signals of similar strength per character (Fig. 1). We did not find a bias in this direction; indeed, there was no significant difference between the cranial and postcranial per character ci values.

We produced a simple linear model expressing partition CI in terms of the log of the number of taxa and the log of the number of characters across all 124 partitions. All terms were highly significant (multiple R^2 = 0.4965, p = 2.2e-16). A subsequent paired t-test of the residual CI values from this model revealed no significant difference (t = -0.5535, p = 0.5820) between cranial and postcranial partitions. We note that other variables have been demonstrated empirically to influence CI (Donoghue & Ree, 2000; Hoyal Cuthill *et al.*, 2010), but our simple model was sufficient to remove the apparent discrepancy between cranial and postcranial CI values in this case.

There was a highly significant difference between cranial and postcranial partition RI values (Wilcoxon test: V = 1613, p = 1.660e-06); more homoplasy in the latter, and the opposite pattern to partition CI and mean per character ci. The RI was only marginally influenced by the log of the number of characters in a simple linear model (multiple R^2 = 0.0290, p = 0.1682), and the residual RIs from this model were highly significantly different too (V = 1615, p = 7.71e-06). Homoplasy Excess Ratio (HER) values were also significantly different (t = 2.042, p = 0.0455), with cranial partitions having a higher mean (0.5675) than postcranial partitions (0.5207); more homoplasy in the latter. However, when the (admittedly non-significant) effects of matrix dimensions were modelled out (multiple R^2

= 0.0042, p = 0.7758), there was no significant difference in residual HER (t = 1.8519, p = 0.0689). Levels of homoplasy in cranial and postcranial partitions appear to be broadly similar, therefore, with differences detected in opposite directions for the partition CI and mean per character ci on one hand, and the RI and HER on the other hand (prior to controlling for matrix dimensions).

We strongly advocate the use of the HER rather than the ensemble CI as an index of homoplasy and data quality. However, we note a possible complication in the calculation of the HER with respect to the distribution of missing values in a matrix. In particular, the procedure does not distinguish between known states (the data) and missing entries (which are *not data*). The effects of missing entries are strongly dependent upon their distribution. A taxon for which all characters code '?' can resolve anywhere in the network with no cost. This will obfuscate the search for MPTs and will obliterate any resolution in the strict consensus. When these same missing entries are randomly redistributed across all taxa, it becomes highly unlikely that individual terminals will bear such a high concentration of '?'s, and resolution becomes more probable. As such, the precise distribution of missing entries may more properly be regarded as an intrinsic property of the matrix (one that should be held constant, along with the frequency distribution of states across characters) rather than as data to be permuted along with the known states. An allied problem is the treatment of additive binary codings and contingent characters. In the former case, the permutation of codes in successive linked columns may yield meaningless combinations, but this is easily overcome with the use of multistate ordered coding. In the latter case, however, the position of an inapplicable code is contingent upon the state of some other character. For example, one character codes for the presence (1) or absence (0) of some feature and the second codes for the form of that feature (coded '?' if the first character codes '0'): permutation again yields meaningless combinations. Although this can be circumvented with the use of more inclusive multistate characters, these do not convey the same information and make different assumptions regarding homology. Blocks of contingent characters are more appropriately regarded as specifying particular types of character state trees. We therefore suggest that the HER permutation step might be modified in two ways; firstly by keeping the positions of missing entries static and secondly by permuting the codes for contingent characters as blocks (c.f. Wilkinson, 2001). These undesirable missing data effects account for occasional estimates of the HER below zero (e.g., Gonzalez-Riga *et al.* 2009; postcranial partition).

We note that the absence of a clear difference between cranial and postcranial

levels of homoplasy does not necessarily imply that *additional* characters of equivalent phylogenetic informativeness can be garnered from the two partitions with comparable ease. For example, a dataset may contain 100 cranial and 100 postcranial characters with identical levels of homoplasy. However, the postcranial characters may have been selected from amongst many (potentially more homoplastic) candidates with enormous care (and represent all the practically extractable data), whereas all *potential* cranial characters could be of uniformly high quality and of much greater abundance. Our conclusions therefore necessarily relate to the *coded* data.

2.5.1 A Significant Minority of Cranial and Postcranial Partitions have Incongruent Signals

Our ILD test results demonstrate that the majority of our 62 sampled data sets (69%) have congruent cranial and postcranial character partitions. However, this leaves 19 data sets in which there is significant incongruence (Allain & Aquesbi, 2008; Anderson *et al.*, 2008; Asher *et al.*, 2005; Asher, 2007; Beck *et al.*, 2008; Ezcurra & Cuny 2007; Friedman *et al.*, 2007; Gaubert *et al.*, 2005; Hill, 2005; Holland & Long, 2009; Li *et al.*, 2007; Matsumoto *et al.*, 2009; Ruta & Coates, 2007; Sanchez-Villagra *et al.*, 2006; Spaulding *et al.*, 2009; Sues & Reisz, 2008; Vallin & Laurin, 2004; Wiens *et al.*, 2005; Worthy, 2009). Assuming a significance level (false positive rate) of 5%, one would expect three or four data sets to be significantly incongruent by chance alone (0.05 x 62 = 3.1). Our results therefore suggest that significant incongruence is being detected across our sample of data sets (binomial test $p < 0.00001$, assuming a 5% false positive error rate). We must stress that we make no inferences concerning the overall quality of individual data sets on the strength of these results. In each case, the original authors analysed all cranial and postcranial characters together; *we* imposed the partitions. One conclusion of this paper is that characters should be sampled from all aspects of morphology rather than focusing upon one region. This was the broad approach taken in the original publications (indeed, we *selected* examples where there was a reasonable balance of cranial and postcranial characters).

It is reasonable to ask whether incongruence is a global phenomenon, or whether it is concentrated in particular taxa (Rodrigo *et al.* 1993). Hence, we explored the effects of removing single taxa in a series of first order taxon jack-knifing ILD tests (see online Appendix 3). These highlighted several terminals with marked effects. For example, the

global incongruence in Ezcurra & Cuny (2007) was attributable to just one taxon, *Dilophosaurus wetherilli.* Removing this species resulted in a p-value of 0.236 (rather than 0.005), whereas removing any of the other 13 terminals still yielded $p < 0.05$. Similarly, the significant ILD result for Allain & Aquesbi (2008) ($p = 0.005$) was entirely contingent upon the inclusion of *Rapetosaurus*. Its removal yielded $p = 0.168$, whereas the exclusion of any of the other 23 taxa resulted in a maximum p-value of 0.015. In Sues & Reisz (2008), incongruence was caused largely by the inclusion of *Scutosaurus* (p increased to 0.719 when it was deleted). In this particular data set, the deletion of three other taxa (Lanthanosuchoidea, *Owenetta kitchingorum* and *Sclerosaurus*) also resulted in $p < 0.05$, although the effect was more marginal. Incongruence in the data of Holland & Long (2009) ($p = 0.008$) largely disappeared with the exclusion of *Eusthenopteron* ($p = 0.578$) and was significantly reduced with the exclusion of *Gogonasus* and *Tiktaalik* ($p = 0.118$ and 0.124 respectively). Finally, the significant ILD incongruence detected in the data set from Sanchez-Villagra *et al.* (2006) was variously contingent upon taxon-sampling (median $p = 0.057$, with a minimum of 0.010 and maximum of 0.303 for 20 single taxon deletion ILD tests). In contrast to these examples, the data sets of Anderson *et al.* (2008), Beck *et al.* (2008), Li *et al.* (2007) and Gaubert *et al.* (2005) remained significantly incongruent no matter which single taxa were deleted (the last of these retained $p = 0.001$ throughout). Time constraints prevented us from applying first order taxon jack-knifing ILD tests to all of the significantly ILD-incongruent data sets in this way.

We also applied the first order taxon jack-knifing ILD to data sets that passed the original partition homogeneity test (i.e., $p > 0.05$). For example, in the borderline case of the data from Pujos *et al.* (2007) (ILD p-value = 0.072), 17 single taxon jack-knifed variants yielded a median p-value of 0.072, with a minimum of 0.005 and maximum of 0.229. Again, although the precise taxon sample influenced the ILD result, the median and mean ILD p-values of first-order taxon jack-knifed variants were similar to those for the entire (all taxon) dataset. This was true in all of the 38 data sets we tested in this manner, suggesting that the ILD test generally offers a robust assessment of congruence despite the marked impact of particular taxa.

2.5.2 Cranial and Postcranial Partitions Imply Significantly Different Relationships in a Significant Minority of Cases

The vast majority of our topological incongruence length difference (TILD) tests reported significant differences for majority rule (>80%) trees of cranial and postcranial partitions. The TILD test, at least as interpreted here, appears very difficult to pass, and we suspect a high rate of Type I errors. Even data sets with comfortably high ILD test p-values (i.e., Beard *et al.,* 2009; Gonzalez-Riga *et al.,* 2009) yielded significant p-values from the TILD.

Results from the variants of our incongruence relationship difference (IRD) test were much more similar (although not identical) to those from the ILD test. For the symmetrical difference distance (Robinson & Foulds, 1981) based upon nearest neighbours (IRD_{NND+SR}) and the majority rule consensus (IRD_{MR+SR}), 37% and 36% of data sets respectively yielded significantly different trees from the two partitions. For the test based upon the maximum agreement subtree distances and majority rule trees ($IRD_{MR+AgD1}$), 21% of data sets yielded significant differences (the nearest neighbour variant, $IRD_{NND+AgD1}$, was not implemented). As with the ILD test, we would expect three or four data sets (5%) to reject our null of homogeneity by chance. Our results using this alternative test of partition homogeneity therefore indicate that a significant minority of cranial and postcranial partitions imply different phylogenies. Of the 19 dataset partitions determined to contain incongruent signals by the ILD test, 12 also implied incongruent relationships ($p < 0.05$) using the IRD_{NND+SR} test, 10 using the IRDMR+SR test and 4 using the $IRD_{MR+AgD1}$ test. Previous studies have suggested that the Robinson Foulds (RF) distance measure may be poor at discriminating between congruent and incongruent partitions. This is because the distribution of RF-distances between randomly generated trees (or between trees generated by a time-homogeneous birth-death model) is highly skewed and has a comparatively narrow range of sensitivity (Koperwas & Walczak, 2011; Lin *et al.,* 2011). However our empirical results indicate that the (IRD_{NND+SR}) test application of this measure not only generates a full-spectrum of values (between 0.910 and 0.01), but also tends to agree with the ILD test assessments of incongruence. We note that faster parsimony programs and more efficient algorithms for calculating tree-to-tree distance metrics (e.g., Goloboff *et al.,* 2008; Pattengale *et al.,* 2007) would increase the computational speed of these tests and allow nearest neighbour distances to be calculated between greater numbers of MPTs. Our cap at 1,000 trees in each partition mean that IRD_{NND+SR} results must be treated with caution (although IRD_{MR+SR} and $IRD_{MR+AgD1}$ results were usually based

upon consensus representations of all MPTs). We also note that there are many other possible tree-to-tree distance metrics that could be implemented in such tests, several of which may have more desirable properties than RF and AgD1 (Lin *et al.,* 2011).

2.5.3 What do these Results Imply for Cladistic Analyses of Morphology?

In studying the evolution of form, it is now relatively common to recognize anatomical modules. These are regions of the body (or suites of landmarks) *within* which morphological changes are strongly correlated through evolutionary time, but *between* which there is significantly less coordination. Different selective forces may operate on these modules, and they may therefore exhibit different evolutionary trends (Mitteroecker & Bookstein, 2007; Klingenberg, 2008). In the context of phylogenetic characters, differing pressures on modules may favour particular patterns of convergence and homoplasy, and therefore suites of characters that imply different trees. The skull of many tetrapod groups has often been regarded as biomechanically and functionally somewhat independent of the rest of the skeleton (Ji *et al.*, 1999; Koski, 2007; Mitteroecker & Bookstein, 2008) hence the difficulty of making many inferences about the one from the other.

Does it matter, therefore, which morphological characters we choose to code when attempting to infer phylogeny? It is difficult to escape the conclusion that it does: whether because the *a priori* omission of characters believed to be analogous or strongly homoplastic is standard practice (although rarely documented explicitly), because alternative data sets for identical sets of taxa often yield radically different trees, because the tree(s) derived from a given data set can alter markedly with the omission or reweighting of characters (most obviously when bootstrapping), or because characters are subject to different selective pressures in different modules. The usual approach in morphological phylogenetics is to combine all available data (Kluge *et al.,* 1989). Although it is acknowledged that the patterns inferred from particular organ systems or suites of characters may be misleading (in the same way and for the same reasons that individual characters may merely introduce homoplasy and noise), it is hoped that the combined analysis of all available characters will allow the true phylogenetic signal to emerge from conflicting local homoplasy. How many characters are enough for the emergence of a reliable signal? The issue, naturally, is one of scale, and concerns the quality of signal

typically encountered in character matrices of the dimensions actually generated in real vertebrate case studies, as well as how these characters are distributed across putative modules. What we *can* say, however, is that in studies of this type, an exclusive focus upon characters of either the cranium or postcranium (at the expense of those of the other partition: e.g., Fitzgerald (2010) (craniodental only) and Mayr & Mourer-Chauvire (2004) (postcranial only)) will significantly influence the resultant optimal tree(s) about one time out of three. This is above the baseline expected for merely sampling a smaller number of characters. We suggest that such a practice is inadvisable, therefore, and strongly advocate garnering character data from all anatomical regions. Certainly, we find little evidence to suggest that cranial or postcranial characters contain different levels of homoplasy (once partition sizes are controlled). Hence, our sample of 62 case studies offers little justification for concentrating on, for example, characters of the skull, because these are believed *a priori* to be of greater value in attempting to infer phylogeny. However, where cranial and postcranial signals conflict, contain similar levels of noise and are of unequal size, then the mean per character consistency (ci) is likely to be lower for the smaller partition (Fig. 1). Such approaches to measuring character quality may give the misleading *impression* of a 'cleaner' phylogenetic signal in the more highly sampled partition. Because systematics (like all science) builds upon prior knowledge, this *may* account for the practice whereby successive studies of certain groups allocate increasingly intensive sampling efforts to particular anatomical regions (as explained by Arratia 2009, and also noted in Joyce & Sterli 2012): the process potentially becomes cyclical and self-reinforcing.

When coding fossils, we may not be able to sample across the same suite of characters that we would employ with extant species (Wiens, 2003a,b; Cobbett *et al.,* 2007). For example, in fossil crocodyliformes, the vast majority of characters necessarily come from the cranium (e.g., Hastings *et al.,* 2011; Puertolas *et al.,* 2011; Cau & Fanti, 2011; Turner & Sertich, 2010; O'Connor *et al.,* 2010) and it is difficult to be confident that we are not merely inferring a 'cranial' tree. The only (and indirect) way to answer this would be to conduct parallel tests upon the closest living representatives of the clade. However, the (quite possibly limited) utility of this approach depends upon the phylogenetic proximity of the extant exemplars, the presumed constancy of selective pressures on putative modules through time and across clades (a big assumption: Hunt, 2008; Frazzetta 2012), and the similarity of the available coded data. Hence, while this may be a sensible approach for crocodiles, it offers little in groups whose closest extant

relatives are ecologically or morphologically very divergent from the fossils (e.g., non-avian dinosaurs, osteostracans), or which are entirely extinct.

A related issue in the context of fossil vertebrates is the preferential preservation of hard part characters (bones rather than muscles or other more volatile tissues). An analogous concern, therefore, is whether skeletal and soft-part characters convey an homogeneous phylogenetic signal (Diogo, 2004, p405-416). If not, this has implications for the manner in which fossil vertebrates are interpreted and analyzed (Sansom *et al.,* 2010) and is an area in particular need of detailed future work.

2.6 Conclusions

- Systematists typically abstract significantly more characters from the cranium than the postcranium. Tests for levels of homoplasy in the cranial and postcranial partitions of our data sets were equivocal, depending upon how homoplasy was assessed. Although postcranial partitions had a significantly *higher* mean ensemble consistency index (CI) than cranial partitions, this difference disappeared when dataset size parameters were modelled out. When all data were analysed simultaneously, mean per character consistency indices (ci) were also higher for postcranial than cranial partitions (although not significantly so). By contrast, the HER (which ameliorates the biasing affects of partition size imbalance using a randomisation test) reported the opposite trend: significantly higher levels of homoplasy in postcranial than cranial partitions.
- Whilst combined analyses using all the evidence available may be optimal for phylogenetic reconstruction, the relative congruence of constituent morphological data partitions has rarely been explored. We are the first to do this with a moderately large sample size (62 data sets), applying a consistent methodology and approach. Cranial and postcranial anatomy implied significantly different phylogenies in about a third of our data-sets. Focus on either source of character data to the exclusion of the other may therefore be inadvisable. We tentatively attribute observed incongruence to the operation of disparate selective pressures acting upon different regions of the body. These yield patterns of homoplasy that may mislead parsimony analyses in one or both cases. More generally, we argue that signals within subsets of morphological data should be examined more

routinely.

- Many different metrics of partition homogeneity and congruence are possible, and we have explored only a subset of them here. While the results from the ILD and the implemented variants of our new IRD test were in broad agreement, there were many detailed differences. These differences may help to illuminate the nature of incongruence by exploring several properties of a matrix (Planet, 2006). Similarly, first order taxon jack-knifing offers a means to localise any incongruence that is detected.
- As measured by variants of our new IRD test, trees derived from cranial and postcranial characters of fishes were more similar to one another than were trees derived from partitions of other vertebrate groups. We tentatively suggest that this may reflect some greater degree of modularity in the cranium of tetrapods relative to fishes. There were no significant differences in ILD results between groups.

Chapter 3: An Updated Examination of the Impact of the Fossil Taxa in Parsimony Analyses of Morphology

3.1 Abstract

In a previously published comparative cladistic analysis (Cobbett *et al.*. 2007), forty-five different cladistic data sets were re-analysed to compare the relative contribution of fossil and extant taxa to the resolution, topology, leaf stability and homoplasy as inferred from parsimony analyses. In this chapter I shall extend this work by using more data sets, newer data sets, data sets containing more characters, more taxa and data sets representing new previously unsampled groups such as plants, echinoids, wasps, chitons, and sea spiders. A quicker, more computationally-efficient pipeline to perform these analyses is introduced; using new technology searches in TNT, tree-to-tree distance comparisons in R, and leaf stability analyses in RogueNaRok (Aberer *et al.* 2013). Comparisons of topological distance based upon mean path difference, in addition to the mean Robinson-Foulds distance are used. On average there are barely any significant differences across all data sets found between fossil and extant taxa in these analyses. Within this, there are however some notable datasets, and notable taxa which we discuss further. Overall though, this further confirms and strengthens the findings reported by Cobbett *et al.* (2007) that fossil taxa are little different in their effect, on average in cladistic analyses of morphology.

3.2 Introduction

Traditionally, palaeontological and neontological approaches to phylogenetic inference were rather separate. It is only in the last 30 years or so that both fossils and extant taxa have regularly been analysed together in cladistic analyses. Fossils offer a unique snapshot view of past evolutionary morphologies along with approximate temporal and environmental information (Adrain *et al.* 2001). They may help break up long branches and provide sources information that our closer to splitting events (Chapter 1), and they allow provide an independent “reality check” on molecular phylogenetic hypothesis (Jenner 2004; Wiens 2004; Smith & Turner 2004).

But the empirical effect of fossil taxa relative to extant taxa across a broad range of taxa has rarely been tested. One robust effort came from Cobbett *et al.* (2007) who examined the difference in relationships when single taxa were deleted (first order jackknifing) from mixed analyses of extant and extinct taxa. By comparing 'whole' dataset statistics (MPTs, CI, RI, relationships, and leaf stability) to the same statistics from taxon-jackknifed variants, Cobbett *et al.* (2007) built a fair test in which we can observe, compare and isolate the properties of fossil and Recent taxa. Through their analyses they demonstrated that in some data sets e.g. Dong (2005) the exclusion of key fossil taxa such as *Ottoia* can have a dramatic affect on the resulting phylogenetic inference – in the case of *Ottoia* exclusion it resulted in a significant loss of resolution as measured by the consensus fork index (CFI; Colless 1980), and a significant increase in MPTs. Even more remarkable is that *Ottoia* is scored 38% missing or inapplicable in Dong's (2005) matrix.

That example was one extreme, but across the entire sample of 45 data sets, fossil taxa and Recent taxa had statistically insignificant differences for most effects tested: homoplasy, number of MPTs, and relationships. Only in the case of leaf stability were there noticeably more data sets in which fossil taxa were significantly less stable than recent taxa. The computational constraints at the time mean't that leaf stability could not be calculated for all the data sets in the sample, so only 36 data sets were assessed for comparative leaf stability.

With the advent of MorphoBank (O'Leary & Kaufman, 2011), and Graeme Lloyd's (2009) online collection of matrices there is now more cladistic data online to choose from for this

type of comparative cladistic analysis. Whilst those archived matrices represent perhaps only 4% of the phylogenetic studies that have been published (Stoltzfus *et al.* 2012) – it certainly provides a good base to start from in an attempt to extend beyond the analyses of Cobbett et al (2007) to a more diverse range of fossil groups and to sample data that is newer, and larger in both taxa and characters. In this chapter I re-analyze data sets such as Legg et al (2012) with over twice as many taxa as the largest data set looked at Cobbett *et al.* (2007), to further examine the impact of fossil taxa in mixed analyses.

3.3 Methods

3.3.1 Finding Appropriate Published Data in the Literature

data sets were sourced either directly from the literature by myself, or sourced from phylogenetic data stores such as MorphoBank (O'Leary & Kaufman, 2011), TreeBASE (Piel *et al.*, 2009) and Graeme Lloyd's collection of matrices (www.graemetlloyd.com).

I also tried the literature search method given in Cobbett *et al.* (2007) but found it to be of limited value. For example, disappointingly few of the 14 articles in the *Systematic Entomology* virtual issue on 'Systematics of Fossil Insects' (2009) contained morphology-based cladistic analyses with enough fossil taxa in them (N > 3) to be of use for this study, despite many of them coming-up in the literature search. Limited access to journal articles as ever was also a significant problem in identifying suitable data sets for inclusion, as rather few of the titles and abstracts made it obvious that the published study contained a cladistic analysis of fossil and extant taxa. A lot of papers were sought that looked promising from the title and abstract but upon receipt of the full text (not always always via quick or easy means) were found to be lacking in usefulness for this particular study and its criteria. In this respect I must especially commend MorphoBank in particular for providing additional metadata that particularly aids the identification of mixed extant/extinct taxon data sets: the 'Taxa' page for each matrix e.g. this one corresponding to Spaulding & Flynn's (2012) analysis of Carnivoramorpha (http://www.morphobank.org/index.php/Projects/Taxa/project_id/367) can in many cases quickly and clearly identify which taxa are extinct and which taxa are extant, enabling easy discovery of data sets appropriate for re-use in this context. However, it must be noted that not all data uploaders have added this metadata (e.g Beutel *et al.*'s (2012) analysis of

Adephaga has only one item of marked-up taxonomic metadata (http://www.morphobank.org/index.php/Projects/Taxa/project_id/814)). Still, it is much appreciated that the authors uploaded their data in it's raw, immediately re-usable format to MorphoBank rather than not at all. The matrices I sourced directly from the literature were much more arduous to reformat back into re-usable data and validate. TreeBASE is ill-equipped and (to be fair) not designed to facilitate searching for proportions of palaeontological taxa. A search for “fossil” in the “All Text” field would seem to be the most valuable approach for this database, and this search currently yields over 200 items.

Due to all the above intricacies of appropriate data discoverability, I acknowledge that my sampling of the literature is thus inadvertently biased towards sampling:

1) studies that have deposited their cladistic data in MorphoBank or TreeBASE or are on Graeme Lloyd's site

2) studies published in journals I had legitimate, immediate full-text access to e.g. those in popular journals, Open Access journals, or where the author(s) have kindly made a freely available and discoverable full-text copy of the work on the internet ('green' Open Access).

3.) studies published more recently, from 2009-2013 that I am naturally more likely to be aware of as I was actively engaged in research during this time.

3.3.2 A new taxon-jackknifing analytical pipeline

Having tried using the DELBAT/DELSUM scripts from Cobbet *et al.* (2007) it became apparent that they were not going to scale well for some of the larger data sets I had collected e.g. Legg *et al.*'s (2012) analysis of arthropod phylogeny which is a matrix of 173 taxa and 580 characters. Nevertheless I tried using them, and soon found additional problems: PAUP* 4.0b10 in particular has a known, reproducible and unsolved bug whereby the AgD1 and AgD tree-to-tree distance algorithms will cause PAUP* to crash if passed certain trees to compare. I have provided example data and scripts in Appendix 1 to demonstrate this bug with a simple comparison of one valid MPT to one other valid MPT that consistently crashes PAUP*. There is also the considerable slowness of doing traditional searches in PAUP* relative to the New Technology search methods (Goloboff 1999; Nixon 1999) now available in TNT.

Thus after I had gone to the TNT workshop in California and received expert one-to-one tutelage in scripting TNT it seemed like I should attempt to create a brand-new modernised analysis pipeline to implement some of the taxon-deletion tests of Cobbett *et al.* (2007). A sketch of the new analytical pipeline is provided in Figure 1

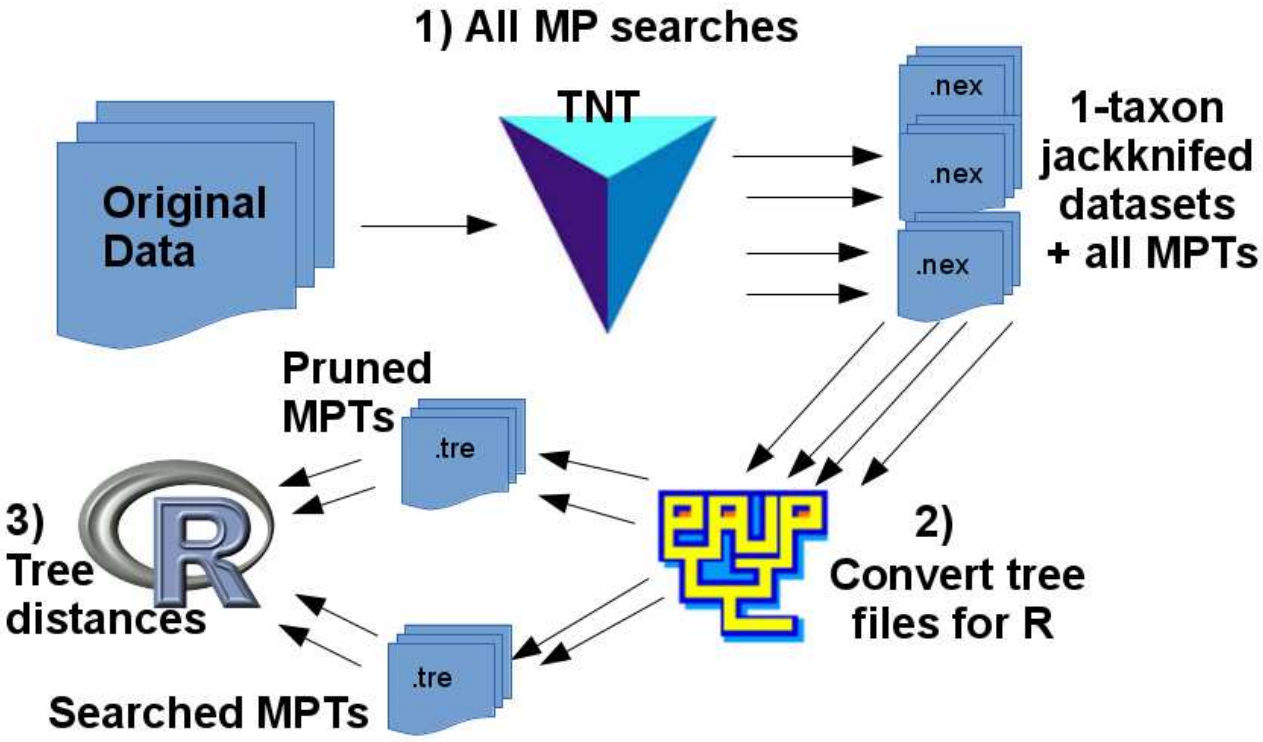


Figure 3.1 Schematic diagram of the workflow that implements taxon deletion comparisons. In step 1 cladistic data matrices are analysed in TNT using New Technology searches (Nixon, 1999; Goloboff, 1999) to find all MPTs for the whole data sets, and all (non-root) single taxon jackknife variations. The results of these analyses, the matrices and the MPTs are exported to PAUP* for step 2 because R cannot directly import either of the tree export formats available natively in TNT. In step two the MPTs from the TNT analyses are simply converted to PHYLIP format with upto 10-character long taxon names. In step 3, the trees are loaded into R using the 'ape' package (Paradis *et al.* 2004) read.tree function and the mean minimum tree distances are found for each set of trees using the treedist function of the 'phangorn' package (Schliep 2011). Full code available online at: https://github.com/rossmounce/extinct_extant_chapter as well as on the CD that comes with the hard copy of this thesis.

This new analytical pipeline of first order taxon-jackknifing uses TNT (Goloboff *et al.*, 2008) to perform all the maximum parsimony analyses using the original assumptions e.g. character ordering & weighting that the original authors used.

The output from R (R Development Team, 2013) from each dataset is a plaintext tab-separated values file (.tsv) of the mean minimum RF (Robinson Foulds 1981) and PD distances (Zaretskii, 1965; Williams & Clifford, 1971) for each deleted taxon, which aids further programmatic access and further command-line manipulation (Figure 1.). To summarize across all the 75 .tsv files programmatically, whilst minimizing human-transcription spreadsheet errors I made use of the 'ddply' function from the package *plyr* (Wickham, 2011) to help automate the process of creating the results summary Table 1.

3.3.3 Leaf Stability Analyses

Since Cobbett *et al*. (2007) significant computational advances have been made in both in tree searching (with TNT) and leaf stability calculation speed (with RogueNaRok; Aberer *et al.* 2013) allowing analyses to be performed on much larger data sets in reasonable times. Software now exists that can perform this type of analysis on trees of 116,334 taxa (Aberer et al 2013). For the leaf stability analyses reported here I used 200 bootstrap replicates, generated from initial analyses using New Technology searches at level 10 – Tree-Drifting, Sectorial searches, Tree-fusing (Nixon, 1999; Goloboff, 1999) on 100 random addition replicates. The 200 standard bootstrap replicates were calculated with 10 random replicates each. RogueNaRok was then used to calculate the leaf stabilities for each taxon. Statistical analyses of leaf support within data sets were performed in R using Mann-Whitney U tests (MWU). Analyses across data sets were performed in R using Wilcoxon matched pair tests, assuming that the null hypothesis is that there is no difference between fossil and Recent taxa.

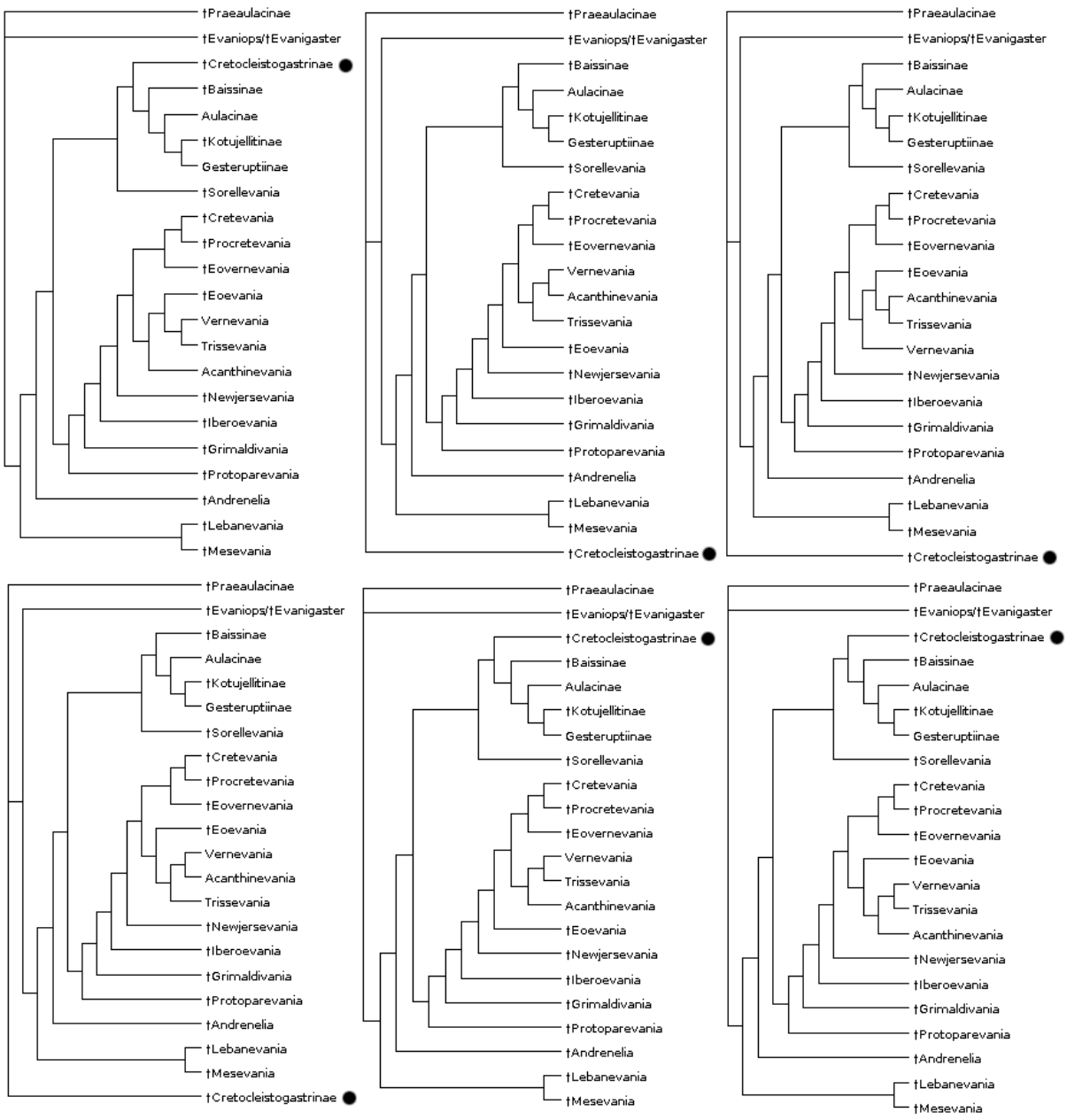


Figure 3.2 Demonstrating a taxon with low leaf-stability. These are all six MPTs from an analysis of the Evanioidea dataset by Penalver et al (2010). The extinct higher taxon Cretocleistogastrinae (indicated with a black circle) appears in two very different positions between the six cladograms. This taxon's leaf stability is 0.340. If one excludes this taxon there is almost no difference between the leaf stability of the fossil (0.926) and extant (0.925) taxa in this dataset, on average. Fossil taxa are indicated by daggers (†).

3.4 Results

All data, code and results are also available in a digitally re-usable form online under an OSI-approved open licence, to maximally enable re-use and re-analysis here on github: https://github.com/rossmounce/extinct_extant_chapter as well as on the CD which accompanies the hard copy deposit of this thesis.

Table 3.1 Summary statistics for the 75 morphological data sets analysed herein from vertebrate, invertebrate and botanical studies. Number of Fossil Taxa is the number of non-outgroup fossil taxa in the data set. Likewise Number of Extant Taxa is the number of non-outgroup extant taxa in the data set. Mean RF, Fossil describes the mean minimum Robinson Foulds distance between the MPT(s) in the original analysis and the MPT(s) from the fossil taxon-jackknifed analyses. Mean RF, Extant is the same but instead for when extant taxa are removed. Mean PD, Fossil describes the mean of all the mean minimum path length distances between the MPT(s) in the original analysis and the MPT(s) from the fossil taxon-jackknifed analyses. Mean PD, Extant is the mean of all the mean minimum path length distances between the MPT(s) from the original analysis and the extant taxon-jackknifed analyses. Mean Difference in MPTs measures the average effect of removing each taxon type on the number of MPTs found – a negative indicates that after removal of that taxon-type less MPTs were found on average, relative to the number of MPTs found in the original analysis. Mean Leaf Stability describes the average leaf stability of that taxon-type in the data set. It ranges from 0 which is unstable, to 1 is stable.

Dataset	**1st Author, Year**	**No. Fossil Taxa**	**No. of Extant Taxa**	**Mean RF**		**Mean PD**		**Mean Diff. MPTs**		**Mean Leaf Stability**	
				Fossil	**Extant**	**Fossil**	**Extant**	**Fossil**	**Extant**	**Fossil**	**Extant**
Acipenseriformes	Hilton, 2009	5	12	2.333	3.621	5.767	9.217	-0.2	2.7	.900	.739
Adephaga	Beutel, 2012	17	30	5.298	7.045	21.413	25.514	49.4	153.7	.996	.970
Amniota	Hill, 2005	59	20	18.499	7.499	79.676	32.852	61.9	25.2	.999	1.0
Archostemata	Beutel, 2008	8	16	6.333	6.219	13.948	15.496	1.3	0.7	.996	.991
Armadillos	Babot, 2012	13	9	8.994	10.183	21.509	23.082	4.1	4.8	.996	.995
Arthropods	Legg, 2012	98	74	7.380	8.053	52.794	49.599	0.39	0.2	1.0	1.0
Avian insectivores	Mayr, 2005	6	12	0.306	1.799	1.164	4.761	-9.8	6.8	.941	.960
Bats	Simmons, 2008	5	23	1.600	1.510	5.875	5.033	1.8	1.6	.992	.984
Carnivora	Spaulding, 2012	35	15	7.968	8.394	39.135	33.594	44.3	207.3	.983	.988
Cetotheriidae	Fordyce, 2013	15	7	2.267	2.524	5.847	7.200	0.8	0.7	.996	.983
Chitons	Sigwart, 2007	26	7	15.153	21.238	40.054	52.821	64.8	119.6	.971	.960
Clawed lobsters	Ahyong, 2006	12	15	4.275	4.542	12.012	13.387	11.3	1.1	.966	.983
Coliiform birds	Ksepka, 2010	11	16	6.887	3.854	18.233	10.077	39	2.8	.974	.973
Crabs	Karasawa, 2006	4	40	7.636	2.292	27.577	12.592	5.0	1.8	.997	.992
Crocodiles	Puertolas, 2011	11	39	3.273	2.886	16.097	14.426	10.5	4.5	.994	.997

Echinoids	Mihaljevic, 2011	6	21	1.238	6.226	4.984	12.211	-0.9	12.2	.957	.979
Ensign scale insects	Vea, 2012	8	38	3.663	3.785	15.215	15.713	680.9	187.5	.988	.986
Euarchonta	Bloch, 2007	16	4	3.971	0.500	8.918	1.785	3.1	0.0	.963	.926
Evanioidea	Penalver, 2010	16	5	4.841	1.800	11.329	4.495	7.4	4.8	.889	.929
Flatfish	Friedman, 2008	5	13	3.133	3.388	7.375	7.637	0.4	3.8	.973	.950
Frogs	Trueb, 2006	12	8	1.435	1.146	4.085	3.574	2.1	0.8	.975	.965
Gnoristinae	Blagoderov, 2004	22	17	13.638	14.310	45.494	45.885	27.5	28.8	.958	.956
Gonorynchiforms	Poyato-Ariza, 2010	17	9	4.177	7.092	11.046	16.272	21.2	99.3	.880	.892
Harvestmen	Garwood, 2011	3	40	5.675	8.066	18.448	24.559	-57.7	-9.2	.870	.978
Horned Crocs	Brochu, 2010	21	15	7.427	7.136	18.555	-61.762	-61.8	-4.5	.968	.966
Junglandaceae	Manos, 2007	5	21	6.871	4.045	16.944	11.040	-10.0	2.7	.963	.958
Kangaroos	Prideaux, 2010	17	17	2.094	2.853	11.221	14.713	0.3	0.5	.996	.990
Lagomorpha	Asher, 2005	38	28	9.312	12.266	36.614	45.862	44.9	64.6	.992	.995
Lepidosauromorpha	Li, 2007	10	23	2.600	2.186	7.922	7.215	0.6	2.8	.959	.971
Mancallinae	Smith, 2011	6	52	7.347	9.910	30.602	44.865	7.0	54.8	.994	.994
Megalyridae wasps	Vilhelmsen, 2010	14	14	1.730	3.428	4.816	11.621	-475.4	-189.6	.595	.564
Mysticetes	Bisconti, 2008	22	12	3.018	2.779	11.454	10.424	1.3	-0.5	.957	.959
Mysticeti	Bouetel, 2006	15	7	1.969	0.371	0.371	1.349	-6.5	-9.1	.975	.971
Neoteleosts	Dietze, 2009	6	10	4.768	3.652	10.456	8.195	3.5	2.9	.961	.958
Odontoceti	Lambert, 2013	21	6	4.584	7.079	12.715	19.018	2.8	22.5	.973	.990
Osteoglossomorpha	Zhang, 2006	19	11	2.604	0.766	8.240	2.386	6.9	2.0	.994	.996
Osteoglossomorpha	GuangHui, 2009	4	11	4.500	5.561	8.088	0.455	0.3	1.5	.981	.962
Pan-Apodiformes	Ksepka, 2013	14	13	4.298	5.769	14.859	18.281	4.3	3.5	.992	.985
Pangolins	Gaudin, 2009	8	8	3.000	2.583	6.315	5.460	0.6	-0.5	.965	.970
Papionin primates	Gilbert, 2013	16	7	8.002	23.357	16.062	43.998	-2.8	-4.0	.957	.940
Pelagornithidae	Mayr, 2011b	5	19	3.100	4.229	8.841	9.985	2.2	1.7	.881	.946
Pelecaniformes	Smith, 2010	8	50	4.565	1.835	20.645	9.614	0.4	1.1	.999	.985
Penguins	Hospitaleche, 2007	5	16	6.300	7.286	12.313	16.371	0.6	2.9	.990	.980
Percomorph fish	Whitlock, 2010	4	22	3.558	2.346	10.972	7.429	2.8	0.8	.875	.838
Pinaceae	Klymiuk, 2012	40	11	65.888	75.571	199.127	229.889	1458.2	2002.7	.826	.996
Pinnipeds	Boessenecker, 2013	17	5	1.739	1.460	6.595	6.366	-7.8	-8.4	.922	.882
Placental Mammals	O'Leary, 2013	39	46	10.195	8.401	45.343	36.985	0.8	1.13	.999	1.0
Placental Mammals	Luo, 2011	59	11	8.360	18.960	37.783	71.396	14.5	18.1	.996	.998
Podocarpaceae	Greenwood, 2013	5	21	6.765	9.183	17.416	21.079	13.0	8.5	.865	.904
Brachyura	Karasawa, 2011	22	14	8.716	8.369	32.442	36.470	2.9	1.1	.996	.998
Polyphaga	Fikacek, 2012	10	29	10.390	15.111	27.863	36.878	549.0	959.0	.910	.950
Porpoises	Lambert, 2008	9	6	3.932	0.167	8.859	0.553	4.9	-0.8	.973	.990
Primates	Beard, 2009	34	4	3.623	3.500	12.608	13.952	1.1	-4.0	.976	.975
Psittaciformes	Mayr, 2010	4	29	18.611	18.460	44.349	46.074	644.2	895.5	.861	.948
Pycnogida	Arango, 2007	4	65	38.272	24.975	173.951	96.312	1697.0	467.1	.937	.990
Ratites	Bourdon, 2009	9	6	0.000	0.000	0.000	0.000	0.0	0.0	.972	.969
Ratites	Worthy, 2012	11	14	1.916	0.143	5.999	0.655	0.3	-0.2	.979	.985
Rhinos	Deng, 2008	26	5	2.123	1.257	7.903	3.483	-22.1	-21.6	.933	.874

Rhizomyinae	López-Antoñanzas, 2013	32	6	10.115	4.551	27.585	14.389	216.3	0.0	.911	.956
Salamanders	Skutchas, 2012	11	10	9.101	6.313	16.210	12.567	12.1	5.7	.941	.920
Sciaroidea	Blagoderov, 2009	7	10	0.714	1.117	2.176	2.791	0.3	-0.2	.926	.961
Sharks	Klug, 2010	11	18	5.742	7.525	15.165	18.164	0.7	16.9	.909	.973
Sharks	Pradel, 2011	15	3	2.513	0.333	6.267	0.943	0.8	0.3	.941	.941
Side-necked turtles	Gaffney, 2011	27	9	2.309	2.900	10.663	13.472	0.4	0.8	.993	.987
Snakes	Apesteguia, 2006	6	12	1.667	4.500	4.387	8.981	-0.5	1.7	.901	.904
Squaliform sharks	Adnet, 2001	7	16	6.896	4.875	18.701	14.127	36.4	2.2	.979	.988
Squamata	Hutchinson, 2012	11	24	1.431	2.349	5.455	6.863	-5.4	15.3	.953	.984
Stem Rollers	Clarke, 2009	5	42	5.510	6.050	22.510	22.405	20.8	69.9	.956	.994
Stingrays	Claeson, 2010	12	27	17.295	10.261	54.495	34.295	1114.5	255.3	.978	.969
Teleosts	Diogo, 2008	5	64	2.333	3.700	11.623	18.355	0.0	9.7	.992	.993
Teleosts	Hurley, 2007	22	6	1.496	1.805	5.995	6.523	-5.7	10.5	.923	.937
Tetrapods	Diogo, 2007	7	73	1.683	4.242	12.416	21.768	18.3	77.4	.999	.999
Turtles	Joyce, 2007	40	22	7.295	7.312	32.970	36.245	57.1	24.0	.973	.977
Wasps	Perrichot, 2009	7	9	1.920	2.796	4.694	6.478	-3.7	-14.0	.856	.861
Waterfowl	Worthy, 2009	10	51	5.224	7.845	18.021	29.677	26.8	61.6	.994	.994

3.4.1 On average Fossil and Extant Taxa have remarkably similar Topological Impact as judged by single-taxon deletions

Across the 75 data sets the mean of the mean minimum RF distances for fossil taxa is 6.58, and 6.77 for extant taxa. The mean of the mean minimum PD distances for fossil taxa is 22.02 and 21.64 for extant taxa. Wilcoxon tests indicated that extant and fossil taxa are not significantly different in either mean minimum RF distance (P = 0.941, 5% significance level) or mean minimum PD distance (P = 0.9985, 5% significance level). Nor was there a difference in the standard deviation of their mean minimum distances, RF (P = 0.5438, 5% significance level), PD (P = 0.6587, 5% significance level).

Subsequent examination of the *within* dataset differences using Mann Whitney U tests show that even where the two group means look different e.g. RF distances this tends to be caused by exceptional individual taxa rather than any general properties of each class of taxa (see data supplied online / on CD; figure 3.3).

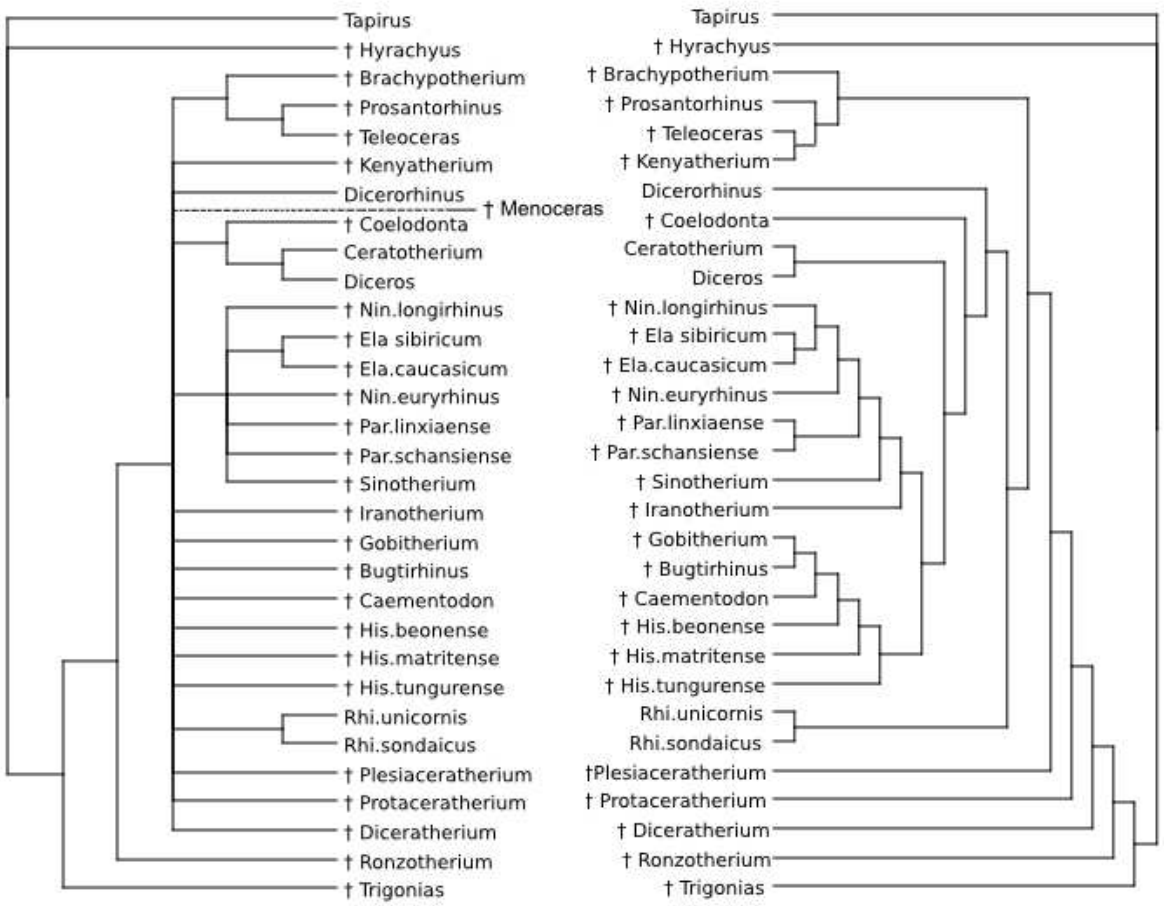


Figure 3.3 Measuring the relationships impact of removing single taxa. Data from Rhinocerotidae (Deng, 2008). The tree on the left is the pruned strict consensus tree. Data set analyzed with *Menoceras* included, yields 37 MPTs. The location of *Menoceras* in the original tree is indicated by the dashed line. The tree on the right is the Searched tree from the same data matrix without *Menoceras* which yields just one MPT. The symmetric difference between these two trees is 6, whilst the path difference is 150.

3.4.2 Fossil Taxa on average do not increase the number of MPTs any more than Extant Taxa

Averaging across all the data sets for the difference in numbers of MPTs upon removal of fossil or extant taxa, there was no significant difference in change in the number of MPTs as assessed with a Wilcoxon matched-pairs test (P = 0.554, 5% significance level), similarly no significant difference between their standard deviations either (P = 0.8252, 5% significance level). There is a strong, and significant positive correlation between the

minimum meanRF and minimum meanPD distances (R=0.96, P = 2.2 x 10^{-16} , Pearson's Product-Moment Correlation Coefficient). Inclusion of fossil taxa demonstrated both negative (figure 3.3) and postive effects on the number of MPTs (figure 3.4).

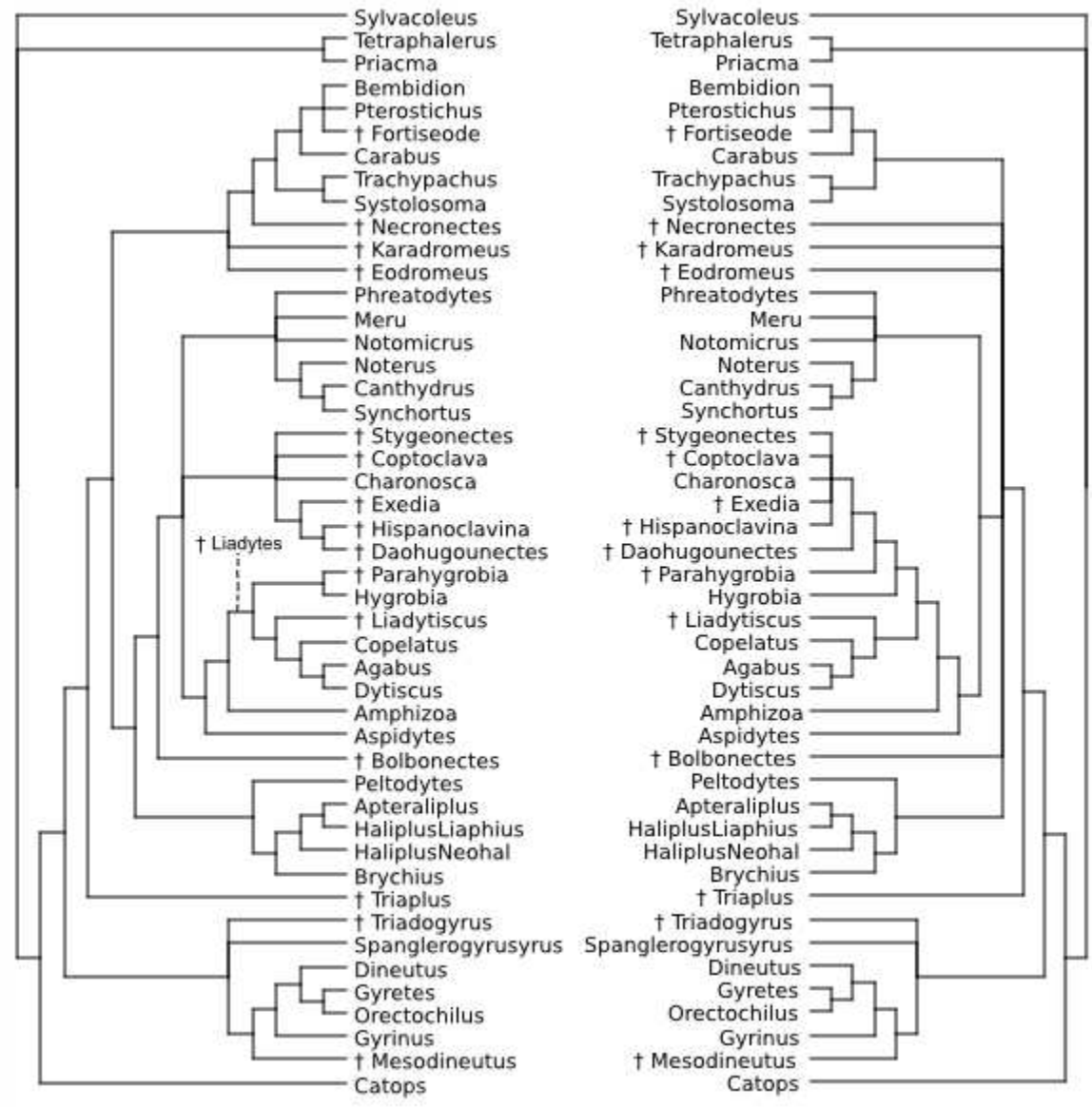


Figure 3.4 A demonstration of a fossil taxon that helps 'support' a phylogeny; in this case the removal of *Liadytes* causes deleterious effects on the resolution (data from Beutel *et al.* 2012). There are less fully resolved nodes on the right (the taxon deleted tree), than on the right (the original whole dataset analysis).

3.4.3 Leaf Stability

30 data sets had a higher mean fossil leaf stability than the mean extant leaf stability. This proportion of trials (30/75) suggests there is no significant difference in leaf stability between fossils and extant taxa (binomial test, P = 0.1053, 5% significance level). It should be noted that the difference in mean leaf stability between extant and fossil was often very slight, with an average absolute difference between the means of just 0.021.

3.4.4 Extreme Taxa

For max.RF (see Appendix 3.2) extant taxa had the highest meanRF distance of all taxa in 41 of the 75 data sets. For max.PD extant taxa had the highest meanPD distance of all taxa in just 37 of the 75 data sets. On an individual taxon-level, the removal of extant taxa caused the highest increase in MPTs in 38 out of the 75 data sets. Using a simple binomial test none of these 'winning' frequencies are significant (P = 0.488, P = 1 and P = 1 respectively).

73 of the data sets had one or more taxa whose deletion increased the number of MPTs relative to an analysis of the 'full' dataset. Only Deng (2008; figure 3.3) and Bourdon (2009) had data whereby the deletion of any single non-root taxon resulted in either the same number or less than the original number of MPTs. 56 data sets had more than half of their non-root taxa cause an increase in MPTs if deleted – a binomial test shows this to be significant ($P = 2.2 \times 10^{-5}$). As a percentage of the total fossil taxa in a dataset, on average 35.8% of fossil taxa caused an increase in MPTs. As a percentage of total extant taxa in a dataset, on average 36.1% of extant taxa caused an increase in MPTs (see Appendix 3.2).

3.5 Discussion

These new results clearly support those first reported by Cobbett *et al.* (2007). Fossil and extant taxa on average have a similar per taxon impact on inferred relationships. These results also support the notion that fossil taxa are no more different in their effect on the resolution of the strict consensus that extant taxa, as also reported by Cobbett *et al.* (2007). However, in this study I obtained a rather different result with respect to leaf stability, than the results of Cobbett *et al.* (2007). In their sample of 36 datasets tested for leaf stability with *RadCon* (Thorley and Page, 2000) they detected a significant difference between fossil and extant taxa on leaf stability, with fossils tending to be significantly more unstable. In this analysis using *RogueNaRok* (Aberer *et al.* 2013) which calculates leaf stability slightly differently – using rooted triplets, I have found no significant difference between extant and fossil taxa for leaf stability although the p-value is quite low (P=0.105).

The biggest advance this chapter offers however, is the code pipeline, which allows any investigator to test their own data sets for these properties. It is sufficiently generalised to accept most data, and uses the common tools that many systematists use, namely PAUP*, TNT and R. The new workflow is much quicker all in all relative to the DELBAT/DELSUM basic scripts of Cobbett *et al.* (2007) and offers easy flexibility and extensibility through the use of R.

3.6 Conclusion

Based on the results presented here and in Cobbett *et al.* (2007), it should now be clear that there is no justification to *a priori* exclude fossils from parsimony analyses of morphology to infer phylogeny. There are a multitude of possible benefits from their inclusion, and multiple different tests available to help one decide *a posteriori* if 'rogue' or 'wildcard' taxa are causing problems in an analysis (Wilkinson 1995b; Thorley and Page 2000; Smith & Dunn, 2008; Pol and Escapa 2009; Mariadassou *et al.* 2012; Aberer *et al.* 2013). There are two avenues of future research I would like to explore arising from this work. The first is a comparative methods review between all the various different tests of taxon stability & inclusion/exclusion criteria. It is unclear to me from the literature which

one of these might be optimal for parsimony analyses of morphology. What are the relative strengths and weaknesses of these methods? Do they all agree for all datasets? If they differ, on what basis do they differ – character evidence or relationship resolution evidence? The second is to extend the current analytical pipeline used in this chapter so that it tests a wider range of tree to tree distance measures. I have already been in contact with the lead author of *TreeCmp* (Bogdanowicz *et al.* 2012) to see how I can integrate his program into my workflow. Once integrated, I would be able to test additional metrics such as matching clusters, nodal splits, quartets, triplets and agreement subtrees.

Chapter 4: A review of the ILD randomisation test: uses and abuses

4.1 Abstract

The incongruence length difference (ILD) randomisation test of Farris *et al.*. (1995a, 1995b) is often applied to systematic data sets comprising qualitatively or logically distinct data partitions (e.g., molecules and morphology, different loci). It has variously been used to assess phylogenetic accuracy (1), 'data combinability' (2), difference in evolutionary history (3) and difference in evolutionary rate (4). Several authors have noted that the ILD test is not designed to address all of these issues, particularly the first two, and that the inferences drawn from test results are often questionable. Here, I quantitatively review the usage of the ILD test in over 250 papers published between 2009 and 2010 - an exhaustive sampling of the papers that I had automated harvesting access to. There appears to be very little consensus on how best to implement the ILD test appropriately. There is often no justification for the number of replications used, or for the search parameters specified. In many cases, the settings used are not reported at all, making it impossible to reproduce the results exactly. Where p-values are not reported, results cannot be re-validated. I conclude with a concise summary re-iterating what is already known about the proper usage of the test and suggest minor extensions that may reveal extra useful information from the test performed.

4.2 Introduction

In 1981, Mickevitch and Farris introduced a new quantitative measure of incongruence of phylogenetic signal between data partitions, which I shall refer to as the 'MF index' to avoid confusion. They were examining the difference in signal between allozyme data and morphometric characters for the phylogeny of *Menidia* and demonstrated that the within-data partition incongruence was greater than the between-data partition incongruence (p 366 – 367). The idea of examining incongruence between data partitions in this manner was then not improved upon until 1991 when Farris further developed the MF index into a proper statistical test with the addition of a randomisation process and an implementation in his *arn* program, presented at the Willi Hennig Society meeting that year (Farris *et al.* 1995a).

For two data partitions, A and B, the MF index is: $L_{AB} - (L_A + L_B)$ where L_{AB} is the optimal tree length in steps of the simultaneous Maximum Parsimony (MP) analysis of both data sources together - the total evidence analysis, L_A is the optimal tree length of an MP analysis of just data A, and likewise L_B is the optimal tree length of an MP analysis of just data B. If the data partitions have low incongruence between each other then $(L_A + L_B)$ will be expected to be only marginally smaller than L_{AB}. Whereas if $(L_A + L_B)$ is much smaller than L_{AB}, then this difference in tree length must be caused by incongruence between A and B. The extra steps required to fit the combined AB data to the optimal tree L_{AB} are referred to as homoplasy and neither the ILD test, nor the MF index help distinguish further what type or cause of incongruence – just that to some extent (or not at all if no difference), that there is some incongruence of signal between the data partitions.

The difference between the MF index and the ILD test, is that the latter is an extension that critically examines the statistical *significance* of the MF index relative to a randomly sampled null distribution of possible MF indices for data partitions of the same matrix parameters and character-state composition as data A and data B. ILD test scores are thus much more comparable between different studies than simple MF indices alone.

4.2.1 Historical Importance and Context

Any sensible discussion and interpretation of the history, development and usage of the ILD test *must* take account of the historical context in which it was developed – there are a number of background factors and themes of which I shall expand upon in this section that are of great relevance to the understanding of the usage of the ILD test.

The ILD test (Farris *et al.* 1995b) was introduced over a decade before the first conference on phylogenomics (in 2006; Philippe & Blanchette, 2007); at a time when the *variety* of molecular sequence data available simply wasn't anywhere near as bountiful as it is today. This was a time when good molecular phylogenetic papers published in respected phylogenetic journals presented data from just one or two loci (e.g. Jacobs et al 1995; Freeman & Zink 1995; Domanico & Phillips 1995; Myers et al 1995; Rosel et al 1995). It was also right in the midst of a long-running controversy and debate around the 'combinability' of different data sources (Swofford, 1991, p327 “To Combine or Not to Combine”; Siddall 1997), sometimes played-out as taxonomic congruence (e.g. Miyamoto & Finch, 1995) “vs” the total evidence approach (Kluge, 1989; Eernisse & Kluge, 1993), and more specialised cases such as “morphology vs molecules” in phylogenetic inference (Wheeler, 1991; Swofford, 1991; Hedges & Maxson, 1996).

Early efforts such as Bull *et al.* (1993; fig. 2) used simulations to provide evidence that combining different data sets in an analysis can provide a 'worse' overall estimate of phylogeny than analyzing data sets separately. Thus many systematists were genuinely unsure of whether to combine or not to combine their data, despite the clear philosophical superiority of using as much relevant evidence as possible – the Total Evidence approach (Kluge, 1989). Even up to the present day both taxonomic congruence and total evidence approaches are still both being actively used, developed and debated: STEM (Kubatko et al, 2009) and *BEAST (Heled & Drummond, 2010) are popular programs that separately infer gene trees to help form a consensus estimate of the overarching species tree (a form of the taxonomic congruence approach). Simultaneously, many papers continue to pursue total evidence (concatenation) methods particularly those that choose to incorporate molecular and morphological data (e.g. Finarelli 2008; Jenner et al 2009; Pretti et al 2009; Schuh et al 2009; Zrzavy et al 2009; Pepato et al 2010; Prevosti 2010; Davis 2010; Fritz et al 2011; Lehtonen et al 2011; Lopardo et al 2011; Carasco et al 2012; Clennett et al 2012; Janssens et al 2012; Ronquist et al 2012; Wood et al 2012)

Thus in recent times the only thing that has changed is the terms by which the debate is framed; instead of taxonomic congruence versus total evidence, debate now centers upon 'coalescence' versus 'concatenation' methods (e.g. sensu Kubatko & Degnan, 2007) which although slightly more specific, clearly has roots in this old debate.

4.2.2 Inappropriate usages of the ILD test

After publication of the ILD test procedure there appeared to be much confusion and uncertainty about what the ILD test could be used for. This led to a multitude of critiques of these various uses (Cunningham 1997b; Yoder *et al.* 2001; Barker & Lutzoni 2002; Dowton & Austin, 2002; Darlu & Lecointre 2002) and led some to question, particularly Barker & Lutzoni (2002) whether the ILD test had any valid use at all.

Farris *et al.* (1995a) contains no discernable suggestion that the ILD test could be used to decide whether to combine or not to combine data – it merely describes the test as a measure of the significance of incongruence between data partitions. Yet after its publication some researchers started to use the ILD test for this very purpose – as a test of “combinability” (e.g. Johnson & Sorensen 1998; Vidal & Lecointre 1998; Carbonne et al 1999; Hoot et al 1999; Spangler and Olmstead 1999), most probably because early versions of the popular PAUP software (versions before 3.1) by Swofford referred to the ILD test as a “combinability” test (Farris 1997; Yoder et al 2001). Other computational implementations of the ILD test such as Siddall's ARNIE and HARDARN, part of the Random Cladistics package (reviewed in Allard et al, 1999b) contain no such assertions about “combinability”. These implementations however, appear to have been much less used than PAUP and thus I speculate that the use of the ILD test to decide “combinability” probably originated from this early PAUP implementation. This incorrect usage of the ILD test as a measure of combinability is discussed and rightly dismissed in Yoder et al (2001) and Barker & Lutzoni (2002). However, neither Yoder et al (2001) nor Barker & Lutzoni (2002) cite Farris' much earlier denouncing of this usage in a conference talk (in 1996) & the subsequently published abstract of that talk (Farris, 1997) in which it is stated that: “incongruence and non-combinability are not the same thing. The idea of testing for non-combinability seems questionable at best”.

Similarly some researchers appear to have initially used the ILD test as a measure of

'phylogenetic accuracy' Cunningham (1997b), or 'topological congruence' (Barker & Lutzoni 2002). The logic behind these uses is difficult to understand given the construction of the test (Mickevitch & Farris 1981; Farris 1995a) – the ILD test is calculated using tree *length* and character congruence NOT tree topology or topological congruence (but see the Topological Incongruence Length Difference [TILD] of Wheeler, 1999 for a measure that does directly test topological congruence). Finally, despite some early criticism (Dowton & Austin 2002), some promise has been shown for using Wheeler's (1999) derivative of the ILD for data exploration (sensu Grant & Kluge 2003), specifically for use in mixed parameter sensitivity analyses (e.g. Sharma *et al.* 2011).

I agree with Hipp *et al.* (2004), *contra* Yoder et al (2001) and *contra* Barker & Lutzoni (2002) that the ILD remains a useful and valid test to assess global partition character congruence and that this is distinct from the inappropriate uses of measuring 'phylogenetic accuracy' or 'topological congruence'. Planet's (2006) comprensive review appears to agree with such usage of the ILD test.

4.2.3 Additional developments

Aside from the many different uses the ILD test was used for; there are a number of other short points about the ILD test that I should note here for the sake of completeness. Sullivan (1996), Cunningham (1997a) and others have suggested that 0.05 may be too liberal a significance threshold to use for the ILD test, hence in the literature sometimes smaller significance thresholds such as 0.01 have been used.

Cunningham (1997b) & Lee (2001) note that invariant and parsimony-uninformative characters should be excluded prior to ILD tests. Lee (2001) demonstrates that such phylogenetically-uniformative characters, if unevenly distributed between partitions can artificially decrease the ILD p-value, sometimes creating false 'significance' – entirely due to the presence of these uninformative characters. My only comment on this is that it is not really much of a “problem” as indicated by Ramirez (2006). Many other cladistic calculations e.g. the ensemble consistency index (CI; Kluge & Farris 1969) require uninformative characters to be excluded and thus this is a standard procedure in most analytical workflows.

Ramirez's (2006) paper contains novel criticisms of the ILD test, for example that localizing exactly where incongruence is with the ILD test is difficult; that the available “strategies are

only feasible for small data sets" - the computational power available now in 2013 makes ILD tests, and taxon-jackknife ILD tests quite feasible even for data sets comprising of many hundreds of terminals. Ramirez (2006) also notes that ILD test results can 'conflict' with partitioned Bremer support results – but this is because the two tests are measuring fundamentally different things; partitioned character support at nodes (local) versus the congruence of entire data partitions (global). These methods should be viewed as complementary and do not give exactly the same information.

Ramirez (2006) figure 2 shows that duplicating taxa lead to undesirable ILD test results. But I question if this is a realistic example. As with uninformative characters, if a cladistic matrix contains uninformative or completely duplicative taxa these are usually removed prior to analysis with 'Safe Taxonomic Reduction' as there is little point including them in the analysis if they include no new phylogenetic information (Wilkinson 1995).

Finally, the main novel critique that Ramirez (2006) presents: "hypercongruence" seems to me to confuse topological congruence and character congruence. Ramirez (2006) figure 4 presents an excellent example of why character congruence and topological congruence are not the same thing. Whilst Ramirez (2006) presents this difference as a problem of the ILD test, I instead suggest interested readers carefully examine the character matrices to rationalize the ILD results therein presented.

Dolphin *et al.* (2000) present an interesting additional procedure that one can choose to perform to help disambiguate between the incongruence due to 'noise' and the 'real' incongruence due to conflicting phylogenetic signals. Despite being well-cited (over 200 times according to Google Scholar), few researchers appear to have chosen to actually implement this additional procedure. Quicke et al (2007) boldly propose that such noise imbalances can be simply 'corrected' for, on the basis of the simulations they run in their paper. However, their arcsine-transformed ILD-metric does not convince me because I do not believe the simulated data matrices they analysed are biologically similar to real molecular or morphological character data. Their method of representing trees as unambiguous binary characters supporting each node, bears little relation to the complexity of empirical character data sets, and furthermore the calculation of RI depends on the number of states allowed per character (Hoyal Cuthill et al 2010) – this is much more variable than just binary in empirical data sets. Thus I do not think their results are necessarily transferrable to real empirical data. Quicke et al (2007) has relatively few citations so far (about 24 according to Google Scholar), and of those almost none of these

actually implement the suggested arcsine-transformed ILD test.

4.2.4 Raison d'etre for this review of ILD test usage

Given the long and complicated history of the ILD test, the various debates and usages of it – I thought it might be instructive to see how people are using the ILD test *now*. Given how many papers all the salient info is spread across one can hardly blame authors for mis-applying or mis-reporting the ILD test, but prior to writing this chapter I had noticed a few odd ones, which partially spurred this review. In doing this review I hope to highlight common errors and misperceptions so that future usage of the ILD test by the research community can be improved, standardised and made more re-usable for comparative cladistic analyses e.g. (Fisher-Reid & Wiens 2011).

4.3 Methods

Literature Search

I performed a literature search for papers published in the year 2009 or 2010, which cite Farris *et al.* (1995b) using the Thompson Reuters Web of Science 'SCI-EXPANDED' database which returned 443 articles. I exported the bibliographic data corresponding to these 443 articles to a bibtex file which is available online for re-use & validation (https://github.com/rossmounce/thesis_ESM/blob/master/ILD_chapter/443_ILD_citing_papers.bib). I then passed Paperpile (http://paperpile.com/) this bibtex file to help automatically harvest and download the corresponding full-text PDF files for as many of the 443 articles as I had legitimate access to through University of Bath journal subscriptions. After this process I had access to corresponding full-text for 278 of the 443 articles. I was thus apparently unable to access 37% (165) of the papers found by this particular literature search. Reasons for my inability to automatically harvest corresponding full-text for some articles included:

- lack of subscription access through my university (e.g. articles from subscription journals such as *Taxon, Journal of Bryology, Auk, Invertebrate Systematics,* etc...)
- poor / incomplete source bibiographic data from Web of Science that Paperpile could not automatically locate a corresponding article for

- Articles from smaller, more unusual publishers for which Paperpile could not automatically harvest full-text articles from

For the purpose of this analysis 278 articles, whilst not the entire sample for the time period assessed, is certainly a large enough sample, from a representative variety of journals (84) and authors from which to draw conclusions from. It was thus not deemed worth manually trying to obtain full-text copies of the remaining 165 papers. The 278 automatically harvested would suffice.

The distribution of the 278 full-text articles across the 84 different journals was remarkably non-random; over a third (101) of the full-text articles came from the journal *Molecular Phylogeny and Evolution*. Given the subject matter of the journal and the volume of articles its publishes (764 over that 2 year period), this distribution is perhaps to be expected.

4.4 Results

See electronic supplementary materials for the full table of evidence and assessment made from each of the 278 papers which the following statistics summarize:

The correlation between citation and usage of the ILD test

Of the 278 papers manually assessed, I could positively discern that an ILD test was performed and reported in 254 (92%) of them. In 17 papers (6%) the ILD test was merely discussed or cited, and explicitly not performed. In 5 papers (2%) the ILD test was mentioned in the methods section, but no further discussion or report of the result(s) of the test(s) was found - it seems the author(s) in these cases may have simply forgotten to report the results of the ILD tests they refer to in their method sections.

Reporting ILD test p-values

Of the 259 papers that refer to the ILD test as if they had performed one or more tests, 207 (80%) reported exact p-values as results of the test(s). 48 papers (18.5%) did not report p-value result(s) of the ILD test(s) performed, mostly instead reporting either that the ILD test results were “significant” or “not significant”. 4 papers provided only information that the p-value was less than or greater than a number, an inexact numerical answer.

Reporting ILD test reps

Replications ('reps') determine in part the robustness of the test applied, thus it is important to report the number of reps.

Of the 259 papers that refer to the ILD test as if they had performed one or more tests, only 131 (51%) reported the number of replications of the ILD test used to determine the null distribution. The number of replications used varied: 10, 100, 200, 500, 1000, 1100, 2000, 5000, or 10000 reps were reportedly used. The majority of papers (77) that reported this parameter used 1000 replications. 14 papers reported using more than 1000 replications, whilst 54 papers reported using less than 1000 replications.

Reporting ILD test search methods

Of the 259 papers that refer to the ILD test as if they had performed one or more tests, 175 papers (68%) clearly specified the search methods used for their ILD tests. The vast majority reported using some form of heuristic random addition sequence + tree bisection reconnection (RAS + TBR) based search using PAUP*, however 5 papers are notable for claiming to have performed their ILD test searches using exhaustive “bandb” searches (Matter 2009; Moyer 2009; Saarma 2009; Meredith 2010; Wei 2010).

Reporting ILD test significance levels used

Of the 259 papers that refer to the ILD test as if they had performed one or more tests, only 61 papers (24%) explicitly state the significance threshold (or critical value) by which they judged the significance/non-significance of the ILD test(s) performed. The most commonly reported significance level used was 0.05, which 35 papers reported. But 26 other papers reported using smaller significance levels in their determination, using either: 0.0001, 0.001, 0.005, or 0.01 as their threshold.

Excluding parsimony-uninformative characters

Of the 259 papers that refer to the ILD test as if they had performed one or more tests, only 32 (12%) papers clearly indicated that parsimony-uninformative characters were removed prior to performing the ILD test. Perhaps a small minority of data sets analysed had no uninformative characters to remove, but this seems an unlikely explanation for all.

“Combinability”

It was much harder to identify this factor conclusively, but as a conservative estimate, of the 259 papers that refer to the ILD test as if they had performed one or more tests, at

least 82 (32%) appeared to have performed the test in order to assess the 'combinability' of their data. As an example of this I quote Bloech *et al.* (2009):

“The combinability of ITS and matK was tested using the Incongruence Length Difference (ILD) test (Farris *et al.*., 1994) implemented as partition-homogeneity test in PAUP...”

4.5 Discussion

Confusion still reigns over what the ILD test should be used for. Even though clearly inappropriate, a third of papers assessed here appear to mistakenly use the ILD test as an arbiter of combinability; a function for which the test was not designed. Several significant papers published subsequently also clearly indicate that it should not be used as such (Farris 1997; Yoder *et al.* 2001; Barker & Lutzoni 2002). Furthermore references to the ILD test as a measure of “phylogenetic accuracy”, “topological congruence” or “compatibility” can also be found in the modern usages sampled. It is unclear why this is so.

Regardless of the reasoning behind *why* the tests are performed, there is also cause for concern with the standard of statistical reporting of the ILD test. As demonstrated, a variety of significance levels are being used. This is defensible given the remarks and suggestions of Cunningham (1997b) about the conservativeness of the test. However, from the standpoint of consistency of interpretation between studies, this is less desirable. For example, Ngamskulrungroj *et al.* (2009) chose a significance level of 0.0001, found a P-value of 0.002 and therefore reported no significant incongruence (in most other papers this would have been judged significant). Fully 76% of papers did not explicitly report the significance level they used and 20% did not report an exact p-value which hampers scrutability and re-use of their results. This tallies with similar such studies in other fields e.g. psychology (Bakker & Wicherts 2011), psychiatry (Berle & Starcevic 2007) and conservation biology (Fidler et al 2006) that also show disappointing reporting standards and inconsistencies of null-hypothesis significance tests across papers.

It would have been good to go through the thousands of papers that cite the ILD test and perform comparative cladistic analyses of the congruence of typical nuclear and mitochondrial genes found, extending and generalizing upon the initial investigation of 13 different data sets by Fisher-Reid and Wiens (2011). But given how many studies in this sample don't report enough of the vital information (exact p-values, the exclusion of

invariant characters, the number of replications used, the search method) it appears that the approach of Fisher-Reid and Wiens (2011) (recalculating new ILD tests themselves) is justified.

Although many implementations of the ILD test allow comparisons between three or more data partitions simultaneously, the single p-value result is difficult to interpret in these cases. Giribet (2010) makes this point well as have others. I would recommend always performing pairwise ILD tests between partitions ($(n^2-n)/2$ tests for n partitions) in order to identify the precise source(s) of the incongruence.

Finally, as reported in Allard *et al.* (1999a), I demonstrated that ILD tests performed on the exactly the same dataset, often appear to give significantly different results between the PAUP* 4.0b10 and TNT (Goloboff *et al.* 2008) implementations. As both are closed source programs I cannot examine the source code to definitively conclude which is 'wrong' or 'right'.

4.6 Conclusion

The ILD test has had a long and tortuous history of development, critique, and usage. Debate has flipped back and forth between positions: useful or useless? Relevant facts and observations about it are scattered across many different papers, some of which are not well cited e.g. (Farris, 1997). Thus as I have demonstrated here the usage of the ILD is sometimes rather inconsistent and confused. In order to help the community better understand the test, and how it should be reported I will conclude with some clear recommendations and clarifications:

- The ILD test (Farris, 1995a) is a measure of the character congruence of data partitions in the context of parsimony analyses.
- It is NOT a measure of, or reliable proxy for; (1) the topological congruence between partitions, (2) phylogenetic accuracy, (3) the “combinability” of partitions, (4) optimising model choice (but see a related derivative of the ILD in Sharma *et al.* 2011).

- The ILD test is not appropriately cited as a justification for combining data from different sources prior to analysis (Siddall 1997). Rather, it is sufficient to invoke Kluge's observation that the strongest test of a hypothesis uses the maximum amount of relevant data (Requirement of Total Evidence; Kluge 1989)
- The ILD test *is* a useful way to investigate the character congruence between data partitions. However, care should be taken not to overinterpret what the result means as there could be a variety of explanations for the level of incongruence found (e.g. heterotachy, noise, 'real' differences in evolutionary signal from hybridization).
- Parsimony-uninformative characters must be excluded before performing the ILD test (Cunningham, 1997b ; Lee 2001).
- Always perform the ILD test with a minimum of 1000 replications to be sure of a robust result, preferably using new technology searches to ensure the minimum length is found.
- Always report the exact p-value obtained, the number of replications, the significance level, the software implementation (e.g. PAUP* 4.0b10), and that you excluded invariant characters prior to testing.
- Preferably, perform *pairwise* tests between each partition as these are easier to interpret than multipartition ILD tests.

Chapter 5: A modification of Archie's Homoplasy Excess Ratio in the presence of missing data

5.1 Abstract

In this short chapter, I demonstrate a problem with Archie's (1989) Homoplasy Excess Ratio measure, that arises when applied to matrices containing missing or inapplicable data. Sparsely-populated matrices are commonly encountered in 'supermatrix' studies, palaeontological studies, and studies that make heavy use of contingent character coding schemes. I proceed to demonstrate a logical solution to the problem as inspired by Wilkinson's use of selective character permutation, maintaining the position of missing/inapplicable states, in his *Phylogenetic Inference by Compatibility Analysis* (PICA) program. I implement this idea using TNT and compare this new modified-HER to other measures of homoplasy.

5.2 Introduction

The Homoplasy Excess Ratio (HER; Archie, 1989) was introduced as an alternative measure of homoplasy for cladistic data matrices. Existing measures of homoplasy such as the consistency index (CI; Kluge & Farris, 1969) are known to have several significant flaws. Central to the calculation of HER is a randomisation procedure that operates by permuting character states within characters for all characters in the matrix, thereby disrupting the phylogenetic signal in the original unpermuted state of the matrix. The HER is thus insensitive to the inclusion or exclusion of parsimony-uninformative characters from a matrix, unlike the CI. A large number of such permuted matrices are then analysed under maximum parsimony in order to obtain a distribution of optimal tree lengths corresponding

to permutations of the matrix. This is similar to the procedure implemented by the permutation tail probability test (PTP; Faith & Cranston 1991). The limitations of the latter test have been rehearsed at length elsewhere (Wills 1999). However, the HER differs fundamentally from the PTP, because it does not use this distribution as the means to test a null. Rather, the mean value for randomly-permuted matrices (MEANNS) is used as an estimate of the expected tree length for matrices of the same dimensions and with identical frequency distributions of states as the original. The formula for it's calculation is given below:

$$HER = (MEANNS - L) / (MEANNS - MINL)$$

where L is the optimal tree length of the original dataset and MINL is the total number of character states in the entire matrix, minus the number of characters – i.e. the minimum possible tree length to explain all the character transformations in the matrix.

A worked example of the HER calculation using 34 unordered, equally-weighted, parsimony-informative, cranial characters from the pterosaur matrix of Andres *et al.* (2010) is demonstrated below. Matrix permutations and Maximum Parsimony optimizations for this chapter were all performed in TNT (Goloboff et al 2008) using New Technology searches (Nixon, 1999; Goloboff, 1999):

```
xperm; xmult=level10; rseed*; /* permute matrix, get tree length, set new seed */
```

The above commands were used to generate 1000 matrix permutation randomisations, the optimal tree lengths of which are plotted in the histogram below. The optimal tree length of the original unpermuted matrix (L) is 59, whilst the mean optimal tree length of 1000 permuted matrices is 99.419 (MEANNS). The minimum possible tree length for a matrix of these dimensions and state frequencies (MINL) is given in TNT by issuing the command `minmax*`. In this case MINL is 38.

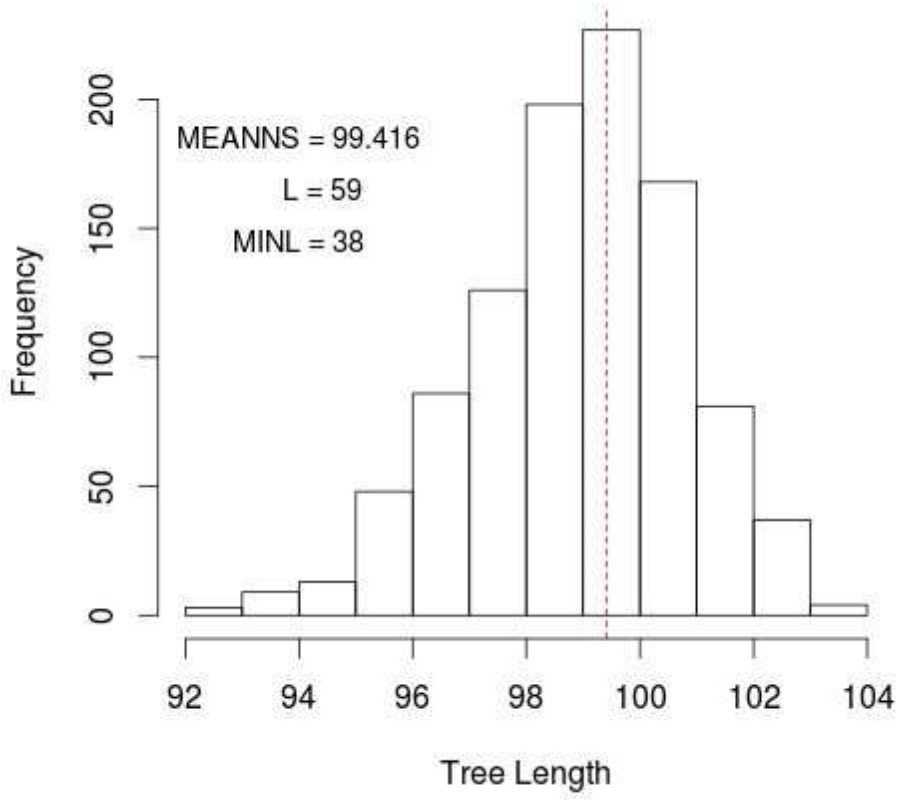


Figure 5.1: The distribution of permuted matrix tree lengths for 1000 permutations of the matrix in appendix 5.1

Thus for the Andres *et al.* (2010) cranial-only matrix;
HER = (MEANNS – L) / (MEANNS – MINL)
HER = (99.419 – 59) / (99.419 – 38).
HER = 0.658

5.2.1 The Problem of Permuting Missing / Inapplicable Data in Matrices

During the course of my research I noticed some curious HER results for some matrices that did not match-up to my expectations. There were some very low <0.1 values, even some negative values coming from the calculation of HER on sparse matrices. Farris (1991, p85) correctly identified that HER could result in negative values but did not precisely describe the circumstances under which this occurs, nor attempt to provide a solution other than using his RI (Farris, 1989) measure instead. Incidentally, Farris was

incorrect in saying that “Negative HER might well be called typical of randomizations” (1991, p85) as I demonstrate in Table 5.1; few empirical cladistic matrices generate negative HER values. I demonstrate this undesirable behaviour with HER on an exaggerated sparsely-populated matrix – figure 5.2. This matrix is composed almost entirely of compatible characters except for character 10 which conflicts with characters 9 & 11 (highlighted in bold). One would therefore expect a relatively low level of homoplasy to be indicated by homoplasy measures.

t1	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1
t2	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	1	1
t3	?	0	0	?	?	?	?	?	?	?	?	?	?	?	?	?	1	1	?
t4	?	?	0	0	?	?	?	?	?	?	?	?	?	?	?	1	1	?	?
t5	?	?	?	0	0	?	?	?	?	?	?	?	?	?	1	1	?	?	?
t6	?	?	?	?	0	0	?	?	?	?	?	?	?	1	1	?	?	?	?
t7	?	?	?	?	?	0	0	?	?	?	?	?	1	1	?	?	?	?	?
t8	?	?	?	?	?	?	0	0	?	?	?	1	1	?	?	?	?	?	?
t9	?	?	?	?	?	?	?	0	**0**	**1**	**1**	1	?	?	?	?	?	?	?
t10	?	?	?	?	?	?	?	?	**0**	**0**	**1**	?	?	?	?	?	?	?	?
t11	?	?	?	?	?	?	?	?	**1**	**1**	**0**	?	?	?	?	?	?	?	?
t12	?	?	?	?	?	?	?	1	**1**	**0**	**0**	0	?	?	?	?	?	?	?
t13	?	?	?	?	?	?	1	1	?	?	?	0	0	?	?	?	?	?	?
t14	?	?	?	?	?	1	1	?	?	?	?	?	0	0	?	?	?	?	?
t15	?	?	?	?	1	1	?	?	?	?	?	?	?	0	0	?	?	?	?
t16	?	?	?	1	1	?	?	?	?	?	?	?	?	?	0	0	?	?	?
t17	?	?	1	1	?	?	?	?	?	?	?	?	?	?	?	0	0	?	?
t18	?	1	1	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	?
t19	1	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0
t20	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0

Figure 5.2: A hypothetical sparse matrix of 19 parsimony-informative characters and 20 taxa.

Indeed the Consistency Index (CI) for this matrix is 0.950. But the HER is -24.

The problem with permuting the entirety of character columns, is that with sparsely-scored matrices such as in figure 5.2, character permutations are likely to artefactually decrease the probability of character conflict, resulting in lowered MEANNS values. Figure 5.3 demonstrates one such permutation whereby the missing states get permuted in such a way that taxon 5 to taxon 20 become completely missing, and two pairs of taxa complete

with data, become identical (taxon 1 is identical in state composition to taxon 2; likewise for taxon 3 and taxon 4).

The HER calculation for this sparse matrix is: HER = (19.04 - 20) / (19.04 – 19), HER = -24.

```
t1      0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
t2      0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
t3      1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
t4      1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
t5      ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t6      ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t7      ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t8      ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t9      ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t10     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t11     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t12     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t13     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t14     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t15     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t16     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t17     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t18     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t19     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
t20     ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?   ?
```

Figure 5.3: A possible permutation of the sparse matrix from figure 5.2

Thus I have identified that HER suffers from a missing data problem which has not been noticed until now. The HER *aims* to be a measure of the 'global' homoplasy across the known states of characters in a cladistic matrix. Yet clearly, the calculation is confounded by the introduction of many *unknown* states (missing data). A desirable measure of global homoplasy should measure *purely* the homoplasy of *known* character states, unaffected by the presence, quantity or distribution of missing data around the known data. Thus I propose here a novel modification to the calculation of HER that estimates MEANNS based-only upon the subset of all possible permutations that keep the unknown states

fixed in the same position as they are in the original unpermuted matrix.

This modified-HER (MHER) reflects the level of homoplasy only in the *known* states of characters, unconfounded by missing data issues, or matrix parameter issues (e.g. the problems of CI; Sanderson and Donoghue 1989; Meier *et al.* 1991; Klassen et al 1991).

5.2.2 Time-Efficient Computational Implementation

Drawing inspiration from the implementation of selective-permutations in Wilkinson's PICA program for compatibility analyses, I reasoned that character column permutations restricted to permuting only information-containing states would help maintain the structure of informative-states in taxa, whilst still allowing permutations within that structure. In this new scheme of permutation, permutations such as that depicted in figure 6.3 would not be allowed. With the help of Pablo Goloboff I have implemented the character permutation of *solely* information-containing states in TNT (Goloboff *et al.* 2008). The script is given in Appendix 6.2. This enables MHER to be calculated for most morphology-based data sets with 1000 replications in a reasonable time (e.g. less than 2 hours to get MHER for Asher et al's (2005) matrix of 223 characters and 68 taxa running on a simple Phenom X3 2.3Ghz desktop computer). Small data sets take just seconds to process.

To make it easier to use this selective-permutation script, I have additionally written a bash script which when passed the name of a dataset in .tnt format, launches TNT, calls the script in appendix 6.2, calculates and reports L, MINL, and the modified-MEANNS value, as well as calculating and reporting the final MHER using *awk* for the final arithmetic operation (appendix 6.3).

The MHER of the sparse matrix in 6.2 is now calculated to give a more intuitive result:

MHER = (25.4815 - 20) / (25.4815 – 19) = 0.845715

Of the informative data states that are in the data matrix, this relatively high value of MHER demonstrates that per character there is relatively little homoplasy in the matrix relative to the optimal cladogram.

5.2.3 How many replications are needed for a robust MEANNS estimate?

Archie (1996, p162) states that “in most cases… with good precision... no more than 25” matrix permutation reps are needed to calculate MEANNS, without supporting evidence. Whilst I agree that there is relatively little variance, I observed it to be non-negligible. Thus here, I quantify the variance using the Carrano & Sampson (2008) dataset. Ten-thousand matrix permutations and their optimal lengths were recorded. The mean optimal length of these ten-thousand permutations was 365.57 (to 2 d.p.), the maximum value was 375, the minimum value was 353 and the standard deviation was 2.96.

Are parametric statistics really justified in this case? Most matrix permutation tree length distributions are significantly non-Normal according to tests such as Anderson-Darling's (data not shown, see also figure 5.1 for a graphical example). But with large sample sizes such tests of normality are often over-powered and give a 'significant' result even if deviation from normality is relatively small, thus I will continue to treat these tree lengths as if they were normally distributed.

5.2.4 Is the HER missing data problem significant in real matrices?

It is all fine and well to demonstrate a problem that could happen *in theory*, but does the HER suffer missing data effects in practice? To assess the difference between HER, MHER and CI, on real empirical matrices from the published literature I tested 60 matrices. Some were garnered directly from the literature (PDFs) by myself, whilst others were gratefully obtained from Graeme Lloyd's excellent shared collection of matrices (http://www.graemetlloyd.com/matr.html), from MorphoBank (O'Leary & Kaufman 2011), or from TreeBASE (Piel *et al.* 2002). The results of these analyses are below in Table 5.1. Most were conducted with 1000 matrix permutation reps, except the 10 molecular supermatrices, which I calculated using only 100 matrix permutations. Admittedly parsimony may not be the best method for inferring phylogeny from molecular sequences but I have included these 10 matrices nonetheless to demonstrate the effect of very sparse matrices upon HER.

5.3 Method

All data sets were analysed in TNT (Goloboff *et al.* 2008) using New Technology searches

(Goloboff, 1999; Nixon 1999) xmult=level10 + bb as well as the scripts provided in the appendices for chapter 5.

5.4 Results

Table 5.1 An empirical comparison of measures of homoplasy, sorted by missing data, for 60 assorted data sets including 7 molecular data supermatrices which are indicated with an asterix*

1stAuthor Year	Group	#Chars	#Taxa	%Miss.	HER	MHER	CI	mean ci
Thomson 2010	Testudines*	3406	213	98.4	-INF	0.787	0.994	1.000
Csiki 2010	Theropoda	364	36	87.6	0.231	0.248	0.423	0.542
Kurochkin 1996	Enantiornithes	122	40	80.7	0.295	0.421	0.692	0.790
Hinchliff 2013	Cyperaceae*	16016	435	78.7	0.778	0.801	0.414	0.859
Wolsan 2010	Carnivora*	9753	52	76.6	0.207	0.229	0.517	0.855
VderLinde 2010	Drosophilidae*	14912	180	73.6	0.429	0.240	0.289	0.811
Springer 2012	Primates*	61199	372	68.6	0.775	0.832	0.358	0.863
Pirie 2008	Danthonioid grasses*	11810	298	68.1	0.846	0.878	0.577	0.940
Moore 2011	Angiosperms*	69513	246	64.6	0.879	0.899	0.214	0.584
Brusatte 2012	Theropoda	233	46	57.6	0.448	0.460	0.441	0.566
McDonald 2010	Iguanodontia	135	67	57.4	0.662	0.674	0.489	0.625
Carballido 2010	Sauropoda	104	19	51.7	0.642	0.659	0.703	0.820
Carrano 2008	Ceratosauria	151	18	48.1	0.712	0.724	0.732	0.829
Friedman 2007	Actinistia	195	39	44.6	0.567	0.573	0.456	0.587
Gonzalez-Riga 2009	Titanosauria	102	23	42	0.478	0.484	0.617	0.751
Young 2009	Crocodylomorpha	166	86	40.7	0.792	0.793	0.447	0.646
Butler 2008	Ornithischia	221	46	39.9	0.619	0.624	0.505	0.649
Gaffney 2009	Bothremydidae	175	47	37.2	0.715	0.718	0.581	0.732
Skutchas 2012	Caudata	72	21	35.3	0.261	0.271	0.497	0.641
Sereno 2008	Carcharodontosaurids	60	9	30.9	0.559	0.565	0.738	0.822
Bloch 2007	Plesiadapiforms	173	21	29.6	0.338	0.341	0.448	0.539
Martinez 2009	Ictidosauria	98	12	29.4	0.317	0.326	0.561	0.683
Hospitaleche 2007	Sphenisciformes	70	29	27	0.521	0.524	0.474	0.598
Simmons 2008	Mormoopidae	202	29	27	0.463	0.465	0.408	0.518
Frobisch 2007	Dicynodontia	100	42	26.8	0.649	0.651	0.470	0.598
Lu 2009	Pterosauria	52	16	25.4	0.496	0.500	0.562	0.685
Anderson 2008	Batrachia	216	54	23.4	0.561	0.418	0.248	0.348
Ezcurra 2007	Coelophysoidea	136	13	22.7	0.591	0.593	0.633	0.714
Matsumoto 2009	Dinosauria	76	16	22	0.517	0.520	0.572	0.644
Asher 2005	Lagomorpha	223	68	18.2	0.467	0.562	0.239	0.326

Holland 2009	Tetrapodomorpha	44	22	17.8	0.429	0.435	0.441	0.557
Asher 2006	Afrotheria	112	23	17.2	0.418	0.421	0.405	0.470
Mueller 2006	Choristodera	90	25	16.8	0.457	0.459	0.415	0.569
Weksler 2006	Oryzomyini	99	54	12.6	0.414	0.414	0.285	0.439
Sanchez-Villagra 2006	Talpidae	157	21	12.4	0.528	0.529	0.452	0.568
Shimada 2005	Lamniformes	55	17	12.4	0.551	0.552	0.525	0.640
Cheng 2012	Eosauropterygia	139	35	12.4	0.548	0.549	0.389	0.491
Smith 2011	Alcidae	223	59	12.4	0.584	0.586	0.211	0.377
Friedman 2008	Pleuronectiformes	58	19	11.8	0.582	0.582	0.507	0.596
Asher 2007	Eutheria	190	53	10.3	0.338	0.338	0.227	0.359
Wiens 2005	Hylidae	140	81	8.3	0.373	0.374	0.184	0.269
Mayr 2011b	Pelagornithidae	87	25	8.3	0.447	0.449	0.377	0.499
Pine 2012	Cricetidae	89	36	8.3	0.256	0.257	0.279	0.378
Worthy 2009	Anseriformes	150	62	8.1	0.429	0.429	0.201	0.297
Whitlock 2011	Diplodocidae	189	27	7.6	0.759	0.767	0.739	0.472
Bourdon 2009	Palaeognathae	129	17	6.9	0.936	0.936	0.876	0.922
Venczel 2008	Caudata	35	15	6.7	0.665	0.666	0.629	0.714
Hill 2005	Amniota	48	18	6.5	0.674	0.674	0.634	0.750
Jouve 2006	Crocodylomorpha	116	64	5.3	0.666	0.667	0.335	0.560
Parenti 2008	Ankylosauria	80	31	5	0.726	0.725	0.595	0.736
Mauricio 2012	Rhynocriptidae	90	38	3.8	0.722	0.722	0.485	0.664
Mayr 2010a	Quercypsitta-Like Birds	93	32	3.7	0.371	0.372	0.343	0.484
Gaubert 2005	Pholidota	329	44	3.4	0.521	0.521	0.323	0.412
Smith 2010	Pelecaniformes	464	53	3.3	0.804	0.804	0.445	0.594
Sparks 2008	Etroplinae	80	25	3	0.912	0.912	0.737	0.853
Mayr 2010b	Caprimulgiform Birds	69	10	2.3	0.570	0.570	0.649	0.762
DePietri 2011	Lari	40	12	1.7	0.459	0.460	0.695	0.800
Li 2007	Squamata	62	10	1.3	0.317	0.317	0.647	0.701
Manegold 2013	Picidae	67	27	0.8	0.803	0.803	0.490	0.696
Hurley 2007	Actinopterygii	31	8	0	0.425	0.425	0.653	0.737

Assuming normality, using the function `power.t.test` in R (R Development Core Team, 2013), it can be shown that at the 1% significance level, for a two-tailed comparison between normal matrix permutation MEANNS (HER) and selective matrix permutation MEANNS (MHER), with 1000 replications each, a significant difference (delta) in sample mean of just 1 step in tree length can be detected with power >0.99.

(Two-sample t test power calculation: n = 1000, delta = 1, sd = 2.96, sig.level = 0.01, alternative = two.sided).

A difference of one step in mean tree length between the two types of permutations for 1000 reps in the case of the Carrano & Sampson (2008) dataset translates into a difference of approximately 0.0013 HER. Thus when comparing between HER & MHER results, even for those results that differ in ratio by only 0.001, this difference is statistically significant. Over half of the data sets tested in Table 5.1 thus display a statistically significant difference between their HER & MHER values.

5.5 Discussion

MHER is a useful measure of the average ('global') homoplasy of a cladistic matrix, comparable between data sets of varying taxa, characters, and percentage completeness. Some claim that such 'global' measures are too crude to be of use; Phillippe *et al.* (1996) once asserted that “A good measure of homoplasy must look at the characters locally and not globally... we suggest that homoplasy should be measured locally” (p 1184). I would have to disagree with this. In the context of comparative cladistic analyses, controlled, standardised 'global' measures such as MHER provide useful information – measures of homoplasy that are local to a particular region of a cladogram cannot be compared between studies that do not contain those same taxa. Depending on the hypothesis to be tested either 'global' or 'local' or both approaches may be of utility.

Meier *et al.* (1991, p77) illustrated an apparent negative correlation between HER and the number of taxa in the matrix based upon an analysis of 27 different matrices. I do not dispute this observation but instead question if matrix parameters are the true *cause* of the correlation they observed. It is not inconceivable that the larger data sets they tested also contained more missing data, with an uneven distribution of it in the matrices. In my analyses (Table 5.1), excluding the supermatrices (n=53), there is a small positive correlation (r=+0.028) between HER & #Taxa but it is not significant (p=0.852), and is perhaps better explained by the percentage and distribution of missing/inapplicable data within matrices. In support of this, Archie (1996, p163) also asserts that HER is not affected by matrix parameters (but no evidence is given).

5.5.1 Why has HER seemingly been ignored for so long?

I identified just 75 papers that mention “Homoplasy Excess Ratio” in a Google Scholar search as of 2013-08-01. Many only mention it in passing. One of these hits was my own conference abstract (Mounce, 2011a). I tentatively suggest three factors that might have led to HER being relatively ignored as a measure of homoplasy since it was devised:

1. The computational complexity of the randomisations might have made it relatively unappealing to calculate, relative to other computationally simpler measures of homoplasy. Goloboff (1991) wrote that HER was “difficult to calculate” and Meier *et al.* (1991) said it was “computationally very expensive and for large analyses impractical”. Hence, undue focus was perhaps shifted to its inferior approximations; REHER and HERM (Archie, 1996). Likewise, Fu & Murphy (1999) later wrote that the “substantial computing time” required was “a major limit of the application of HER” back then.
2. Farris's (1991) paper strongly criticizes HER and argues throughout that RI could and should be used instead. Similarly Meier *et al.* (1991) observed an apparent negative correlation of HER with matrix parameters. Perhaps these papers deterred people?
3. Confusion between HER and HERM. As HERM was correctly shown to be identical to Farris' RI (Farris 1989; 1991) it is possible that some did not see the distinction between HERM & HER and thus discounted both when HERM was shown to be redundant. A putative example of this is given by Egan (2006) who mentioned only HERM in her discussion of “goodness of fit metrics”.

A new generation of academics are once again investigating important questions about homoplasy (e.g. Hoyal Cuthill *et al.*. 2010). I have no doubt that there are many further important studies to be done in this area. In future work, I will compare MHER with Goloboff's Data Decisiveness (1991) & Wilkinson's (1997) incompatibility excess ratio's – two other under-utilised descriptive statistics for cladistics.

Chapter 6: Optimal search strategies for *finding* phylogenetic data

6.1 Abstract

In the preceding chapters, I have repurposed data from published studies in order to perform comparative cladistic analyses and thereby to test hypotheses that single cladistic dataset studies alone could not adequately address. In order to re-use these data, I had to *find* relevant studies and *transform* the published data back into a digitally-usable form – neither of which are trivial tasks. In this chapter I explore and critically compare methods of digital literature search in the context of the task of finding morphology-based phylogenetic analyses. Traditional title-keyword-abstract searches with *Web of Knowledge* & *Scopus*, and modern full-text searches as provided by publishers like PLOS & BMC, as well as offline local desktop full-text searching are compared and discussed. I demonstrate that the popular traditional methods are significantly poorer at finding phylogenies (in terms of recall to known relevant papers) when compared to full-text search methods. I conclude that despite three and half years of looking for morphology-based phylogenetic analyses, I can only put a conservative minimum-bound estimate on the number of morphology-based phylogenetic studies that have been published in the 21st century because of inadequate bulk full-text access to journal literature.

6.2 Introduction

Forty years ago a thorough literature search necessitated a trip to a library so that researchers could systematically examine all relevant journals and books page-by-page to scan for the desired concepts and items of interest. More recently, the ubiquitous electronic publication of research on the Internet has enabled less-manual, more computationally expedited methods of literature search using computer software to scan articles and books for relevant terms and concepts in text-form. To help academics find relevant content (and to make a profit by charging for this commercial service) Thomson Reuters released the first version of *Web of Knowledge* (WoK) a 'research platform' for academic content discovery over a decade ago – it launched in 2002 (Thomson Reuters, 2013). Shortly afterwards, Elsevier launched a rival profit-making commercial service called *Scopus* (Fingerman, 2004). Both of these abstract & citation indexing services are now widely used by researchers in non-biomedical biological sciences. One of them; WoK *only* indexes the title, abstract, keywords and citations for each article or book chapter.

It is important to note here that I will not discuss PubMedCentral (PMC) as commonly used by biomedical researchers because on the whole it *only* indexes biomedical content – whilst morphology-based phylogenetic content can occasionally be found indexed in PMC, particularly if it appears in a 'general' journal like *Nature*, *Science* or *PLOS ONE* e.g. Eddy & Clarke (2011), most non-biomedical subject-specific journals e.g. *Zootaxa*, *Palaeontology*, and *Journal of Vertebrate Paleontology*... are not indexed in PMC. Thus PMC cannot be relied-upon for literature searches for non-biomedically relevant topics. As a further demonstration of this, if one searches for 'Zootaxa' in PMC, one can only find four articles from the journal *Zootaxa* in PMC that have been self-deposited by their authors as 'author manuscripts' (context: *Zootaxa* has published over 12,500 articles as of 2013-07-01).

Other relevant digital search services that academics sometimes use include *Google Scholar* (GS; http://scholar.google.com/), *Scirus* (http://www.scirus.com/), *Mendeley Search* (MS; http://www.mendeley.com/research-papers/search/) and *Microsoft Academic Search beta* (MAS; http://academic.research.microsoft.com/). GS first launched nearly ten

years ago as beta in 2004 (Google, 2013). GS can notably achieve 100% recall for some searches (Gehanno *et al.* 2013) and is thus often better than *Scopus* & WoK's recall (e.g. Beckmann & von Wehrden 2012). But the precision of GS is often very poor (Garcia-Perez 2012), since it searches across a much wider body of grey literature: including some blogs, newsletters and non-peer reviewed material It also offers relatively few features with which to constrain or filter searches (other than simple 'by year/journal/author'). Moreover, there is no easy mechanism provided by which hundreds of search results can be exported in a standard format (e.g. bibtex). Thus some have pointed out before that GS is not useful for performing *systematic* literature searches (Giustini, 2013). *Scirus* (also known as Sciverse Hub: they are different interfaces to the same index [California Digital Library, 2013]) allows full text search of a limited subset of the research literature, as well as abstract-only search, and grey literature 'scientific web' searches.

For the purposes of this chapter, when referring to *Scirus* I shall only be referring to the full text search subset of the capabilities of *Scirus*. MS is a relatively new academic search provider and claims to search across a crowd-sourced database of nearly 100 million documents (Mendeley, 2013). MAS is yet another new academic search provider and is still in active development, the service is self-described on their About page (Microsoft, 2013). GS, Scirus and an earlier version of Microsoft Academic Search have previously been compared (Ford & O'Hara, 2008) for searches in 2006 during which GS retrieved the most citations, however the aim and methodology of that study is different to the one presented herein, and I anticipate that all of the databases have improved in performance since 2006.

6.2.1 Preliminary Investigation

Table 6.1 illustrates some of the complexity of digital literature searches. GS finds the most content overall for a simple keyword search over a defined period (2000-2012) but this includes many books and non-peer reviewed grey literature pieces that one would perhaps want to exclude (with no easy way provided to filter these out). GS also notably does not handle wildcards, so I could not perform a more conservative search for phylogeny-related articles with 'phylog*' to catch the words 'phylogeography', 'phylogram', 'phylogenies' etc... For general searches, in agreement with the findings of Chadegani *et al.* (2013), *Scopus* appeared to find more content than WoK but in specific cases (e.g. the searches for Winclada:Table 6.1) WoK occasionally appears to outperform *Scopus* in raw hits.

Table 6.1 A comparison of simple search results for the term "phylogeny" and "winclada" in scholarly documents published between the years 2000 – 2012 inclusive, all searches performed 2013-07-15.

Search Service	Term	Exact Repeatable Search Terms Used	Hits
Google Scholar	phylogeny	http://scholar.google.co.uk/scholar?as_q=phylogeny&as_epq=&as_oq=&as_eq=&as_occt=any&as_sauthors=&as_publication=&as_ylo=2000&as_yhi=2012&btnG=&hl=en&as_sdt=1%2C5&as_vis=1 (excluding citations, excluding patents)	~220,000
Scirus	phylogeny	http://www.scirus.com/srsapp/search?sort=0&t=all&q=phylogeny&cn=all&co=AND&t=all&q=&cn=all&g=a&fdt=2000&tdt=2012&dt=fta&ff=all&ds=jnl&sa=agr&sa=geo&sa=env&sa=life&sa=med&sa=neuro&sa=pharma (articles only, exc. non-journal sources)	43,836
Scirus	phylog*	http://www.scirus.com/srsapp/search?sort=0&t=all&q=phylog*&cn=all&co=AND&t=all&q=&cn=all&g=a&fdt=0&tdt=2014&dt=fta&ff=all&ds=jnl&sa=agr&sa=geo&sa=env&sa=life&sa=med&sa=neuro&sa=pharma (articles only, exc. non-journal sources)	163,015
WoK	phylogeny	Topic=(phylogeny) Timespan=2000-2012. Databases=SCI-EXPANDED	44,946
WoK	phylog*	Topic=(phylog*) Timespan=2000-2012. Databases=SCI-EXPANDED	120,078
Scopus	phylogeny	TITLE-ABS-KEY(phylogeny) AND SUBJAREA(mult OR agri OR bioc OR immu OR neur OR phar OR mult OR medi OR nurs OR vete OR dent OR heal) AND PUBYEAR > 1999 AND PUBYEAR < 2013	127,991
Scopus	phylog*	TITLE-ABS-KEY(phylog*) AND SUBJAREA(mult OR agri OR bioc OR immu OR neur OR phar OR mult OR medi OR nurs OR vete OR dent OR heal) AND PUBYEAR > 1999 AND PUBYEAR < 2013	156,583
Google Scholar	winclada	http://scholar.google.co.uk/scholar?q=winclada&hl=en&as_sdt=1%2C5&as_vis=1&as_ylo=2000&as_yhi=2012 (excluding citations, excluding patents)	~1,840
Scirus	winclada	http://www.scirus.com/srsapp/search?sort=0&t=all&q=winclada&cn=all&co=AND&t=all&q=&cn=all&g=a&fdt=2000&tdt=2012&dt=fta&ff=all&ds=jnl&sa=agr&sa=geo&sa=env&sa=life&sa=med&sa=neuro&sa=pharma (articles only, exc. non-journal sources)	292
WoK	winclada	Topic=(winclada) Timespan=2000-2012. Databases=SCI-EXPANDED	18
Scopus	winclada	TITLE-ABS-KEY(winclada) AND SUBJAREA(mult OR agri OR bioc OR immu OR neur OR phar OR mult OR medi OR nurs OR vete OR dent OR heal) AND PUBYEAR > 1999 AND PUBYEAR < 2013	12

However, relative to GS's fulltext search capability, WoK and *Scopus* have a markedly diminished ability to find methodological details that are not usually discussed in abstracts. This is why I am doing this chapter – I have noticed during my research that it is very difficult to accurately search for certain types of phylogenetic methods detail. I will use this

chapter to objectively test the performance of very search methods for finding papers with phylogenetic analyses in them.

It is known that *Scopus* and WoK have non-overlapping journal coverage (e.g. see Figure 1 of Chadegani *et al*. 2013). However, even in cases where journal coverage should overlap, the recall between *Scopus* and WoK varies, as can be seen in Table 6.1. Relative to the *Scopus* search, WoK has an additional six hits from articles in these journals: *Acta Paleontologia Polonica*, *Apidologie*, *Coleopterists Bulletin*, *Geodiversitas*, *Journal of Biogeography*, and *Journal of Vertebrate Paleontology* (full article bibliographic data for these *Scopus* and WoK 'winclada' searches are provided online (Mounce 2013c,d) and on the CD provided with the hard copy of this thesis. These six missing articles do contain the word 'winclada' in the abstract and all six journals are indexed in *Scopus* according to their list of indexed titles as of April 2013 (Sciverse, 2013). It is unclear why *Scopus* cannot find the word winclada in these six article abstracts and thus does not return them in its results, which it *should* be able to find. Thus for the winclada search *Scopus* has a recall of at best 66% (12/18) for abstract search, assuming perhaps unsafely that WoK finds *all* the relevant article abstracts that it indexes (?unknown). *Scirus* finds 16% of what GS finds for the winclada search and this perhaps reflects the narrowness of the biological journal sources that *Scirus* restricts itself to – just ten relevant to biology: Elsevier, Wiley, Springer, BMC, the PMC OA full-text subset (which includes all PLOS journals), OUP, CUP, NPG, Royal Society & Hindawi.

6.2.2 Scirus does not cover many important natural history journals

This restricted range of sources used by *Scirus* search makes it unsuitable for phylogeny-related searches. Relative to the range of sources searched by WoK (albeit abstract-only in this database), I find that at the article level – the unit that counts in these matters – the ten full-text sources searched by *Scirus,* listed above account for only ~61% of nearly 10,000 phylogeny-related articles published in the year 2010, found and classified by publisher from a search of WoK that I performed as part of an international collaboration (Stoltzfus *et al.* 2012, detailed source data for this estimate in Mounce 2013e]. Non-biomedical phylogenetic content included in WoK but likely to be outside of the *Scirus* full-text search capability includes: AAAS (*Science*), Academica Sinica (*Botanical Studies*,

Zoological Studies), Allen Press (*Herpetologica*, *Journal of Mammology*, *Phycologia*), AMNH press (*Bulletin of the AMNH*, *AMNH Novitates*), ASIH (*Copeia*), ASPT (*Systematic Botany*), CSIRO journals, Magnolia Press (*Zootaxa*, *Phytotaxa*), Taylor & Francis (*Journal of Vertebrate Palaeontology*, *Journal of Systematic Palaeontology*, *Journal of Natural History*). Thus *Scirus* lacks coverage of many key journals for morphology-based phylogenetic analyses.

6.2.3 The long-tail of phylogenetic content

The distribution of 21st century phylogeny-related articles across publishers and journals is remarkably long-tailed. Surprisingly, using the methods of Stoltzfus *et al.* (2012) one can infer that *Zootaxa* (a primarily taxonomic journal not especially known for phylogenetics), might now be the 6th largest container of 21st century phylogeny, just behind PLoS ONE (as shown in figure 6.1).

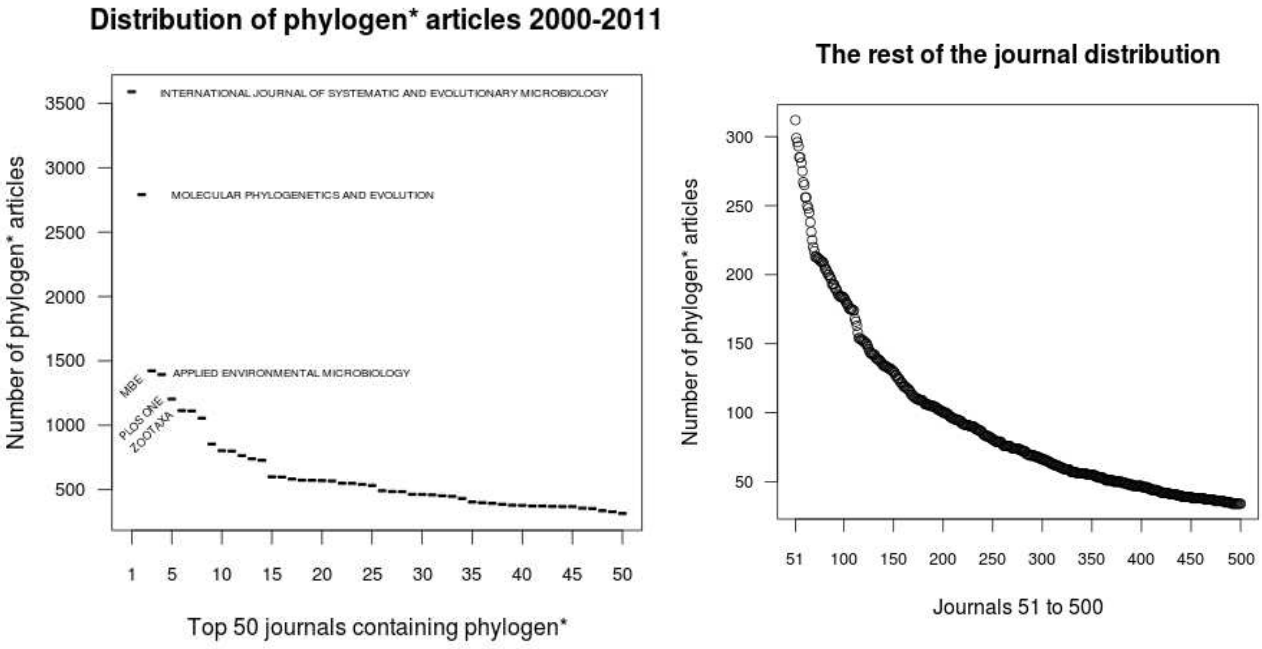


Figure 6.1 Illustrating the long-tail distribution of phylogeny-related articles across hundreds of different journal titles indexed in WoK, from 2000-2011. Full corresponding data is supplied in the supplementary materials and online (Mounce 2013f).

There are two important things that figure 6.1 highlights:

- Publishing trends have changed. Now a significant proportion of academics are choosing to publish in megajournals like PLOS ONE and Zootaxa instead of low-volume traditional journals (this is no bad thing in my opinion).
- Aside from those 6 or 7 journals estimated to be publishing a lot of phylogenetic content – the distribution of the thousands of other phylogenetic articles published each year is remarkably long-tailed. Phylogenetic articles are scattered across at least 1000 journals at a minimum estimate (Stoltzfus *et al.* 2012 data)

6.3 Methods

In order to rigorously examine the capabilities of traditional title-abstract-keyword databases WoK & *Scopus* for finding morphology-based phylogenetic analyses published this century, I compared their precision and recall to three different defined corpora of journal articles to which I have full text local desktop access:

A) 'Zootaxa'. The entire set of articles published in the journal Zootaxa from 2001 up to Issue 3690 (1) [11th June 2013] inclusive, consisting of 12490 PDF files downloaded direct from the publisher website: http://mapress.com/zootaxa/ . This set notably includes both large monographs and small erratum notices (see Figure 6.2 overleaf).

B) 'PLoS'. All articles published across seven PLoS journals: One, Biology, Computational Biology, Genetics, Medicine, Neglected Tropical Diseases, and Pathogens from 2003 to 2010-06-04, consisting of 20694 articles obtained via BioTorrents (Langille & Eisen, 2010).

C) 'BMC'. A complete set of 7948 open access articles containing the stemword 'phylogen*' from 166 journals that BioMed Central publish (2000-2011, full details of each and every article are provided [Mounce, 2013a])

The corresponding PDF's of each corpus were placed in separate self-contained folders. Within these folders I used basic unix command-line tools e.g. *pdftotext* & *grep* searches to identify articles that are very likely to contain relevant content, and then manually scan-read the articles myself to classify each search hit. Thus I can be very confident of the *actual* word content of all the articles in these corpora, relative to the search results provided by the various academic search providers.

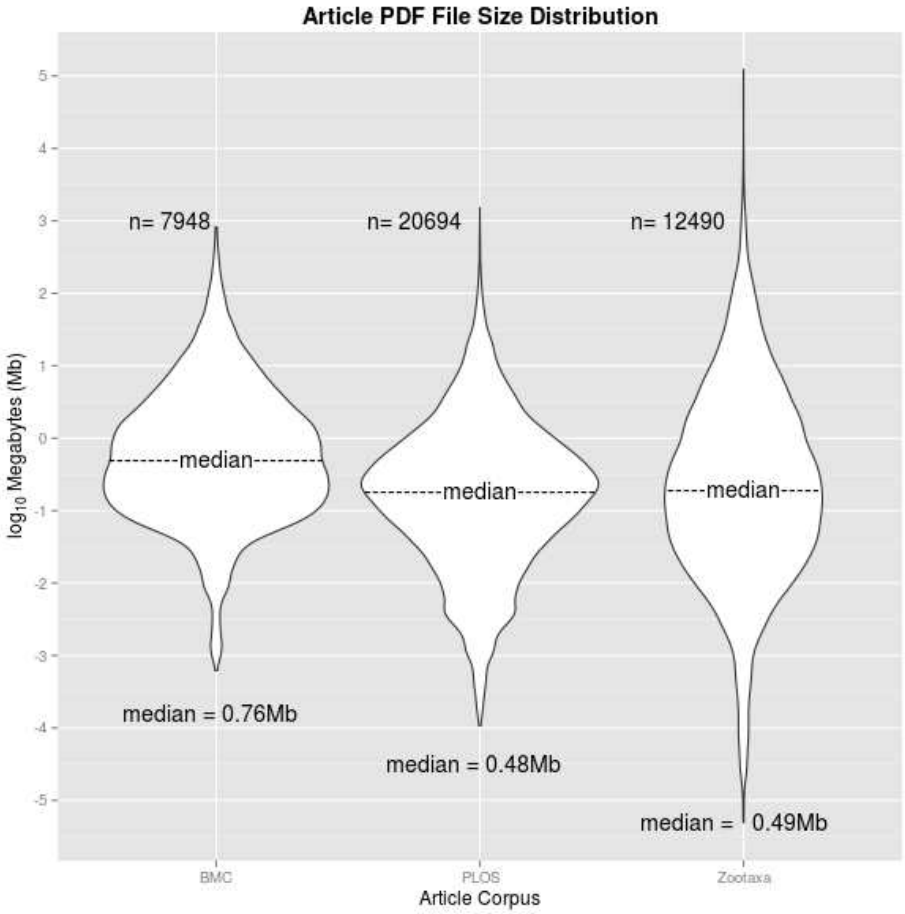


Figure 6.2 Article PDF file size distribution of all articles in each of the corpora. Mean file size across all three corpora is almost exactly 1MB (Mounce 2013b).

6.3.1 Testing the computational efficiency of local full-text search options

I speed-tested three different command-line methods to determine which search methods were the quickest in terms of real time. Table 6.2 (overleaf) shows that of the methods that work from 'cold' - spawning embarrassingly parallel grep processes on each of the three cores available returned results in the quickest overall real time. GNU parallel [21] did not speed-up the process because the I/O rate of hard drive access was the limiting factor and this is limited to single-thread access [22]. If one scaled-up this experiment and searched across multiple different disk drives, or used solid state drives (SSD's) GNU parallel may well scale better. Using *Recoll* [23] returns search results across the entire corpus near instantaneously. But it requires many hours to index all the article content beforehand before searches are performed, so is not truly comparable. *Recoll* indexing only needs to be done once however to gain this speed benefit for *all* searches and comes with the

advantage of a simple GUI interface to perform searches and explore the results.

Table 6.2: A demonstration of different local desktop full text search methods, as performed on the Zootaxa corpus, searching for the term 'winclada' on a simple Phenom X3 2.3Ghz desktop computer, with all files placed on a single HDD.

Method name		Real time (s)	User time (s)	Sys time (s)
'Parallel grep'	time find . -name "*.txt" -print0 | xargs -0 -n 1000 -P 3 grep -iRl "winclada" > out.tzt	15.989	44.163	0.384
'Single process grep'	time grep -iRl 'winclada' . > out.tzt	42.993	42.803	0.180
using GNU parallel	find . -type f | time parallel -j+0 'grep -il winclada {}' > out.tzt	61.11	34.47	0.481
using *Recoll*	(after indexing all the PDFs overnight)	< 1	n/a	n/a

6.3.2 Justification of search terms used, and the use-case

Phylogenetic analyses are computationally complex and thus almost always involve the use of a specialised computer program, of which there are few in existence: e.g. PAUP* (Swofford 2002), Winclada (Nixon 2002), NONA (Goloboff, 1999b), TNT (Goloboff et al 2008), Hennig86, MEGA, POY, PHYLIP, Phyml, Garli, RAxML, MrBayes, PSODA. A well-reported paper thus *must* make reference to one or more of these programs in the full text of the work – this offers a nicely constrained starting point for searches. For this chapter I ignored POY, Garli, MEGA, RAxML, Phyml, and MrBayes as they are never or rarely used in morphology-based analyses. PSODA (Carroll *et al.* 2007; Carroll *et al.* 2009) seems to have only been cited 3 times according to GS and thus can also be safely discarded. That leaves just six terms and their variants to search for in the quest for morphology-phylogenies: PAUP*, Winclada, NONA, TNT, Hennig86, and PHYLIP.

I proceeded to use those same search terms (and more) to try and re-find the identified content in the three publisher-delimited corpora using many different academic search providers, benchmarking for recall against the *known* location of the searched-for terms, in articles, in the corpora.

6.3.3 Distinguishing between morphology-based & molecular-only analyses

Some of the literature search complexity for morphology-based phylogenetic articles is caused by linguistic complications. Occasionally papers exclusively refer to such morphology-based analyses as cladistic analyses (this terminology is entirely correct, but it makes it difficult to find the paper if one is only searching for the word 'phylogeny' and its variants). Far from being a theoretical occurrence there are many easily documented occurrences of this in the Zootaxa corpus. One can automatically identify papers which mention 'cladist*' and NOT 'phylog*' with this simple grep:
find . -name ".txt" -print0 | grep -iR -l 'cladist' | xargs -n 1 grep -iL 'phylog'* .
There are many papers which merely contain one but not the other stem-word (and no cladistic/phylogenetic analysis), but also more seriously there are papers such as Dimitrov & Ribera (2005) in which a morphology-based cladistic analysis is reported and nowhere in the paper; not in the title, keywords, abstract, main text, figure captions or reference list does any variant of the stem-word phylog* occur – quite an achievement really. The inverse condition is also true, there are many more papers in which morphology-based phylogenetic analyses are performed with no mention anywhere in the paper of either cladi* or cladogra* (at least 47 different papers in the Zootaxa corpus, in which morphology-based phylogenetic analyses are performed). This further justifies the approach of trying to find analyses by other means (e.g. the name of the software used).

For the purpose of this chapter “relevant content” and “morphology-based phylogenetic analyses” are defined as papers in which new (biological) morphological-character-based phylogenetic analyses are both calculated and reported in the paper – hypothesised figures (only), or reproduction of a phylogenetic tree(s) from previous studies, or 'supertree' analyses, or phenetic analyses (e.g. UPGMA) of continuous measurements do not count. Analyses that combine morphological and molecular data to build the phylogeny are included in the count. Molecular (only) phylogenetic analyses are not defined as relevant content for the purpose of this exercise, which makes the task more difficult since molecular and morphology-based analyses often use the same methods, software and linguistic reporting style. Moreover there are far more published molecular-only papers which hampers the precision of any literature search for morphology-only phylogenetic papers.

6.4 Results

Based on the initial grep for just 'paup' it was determined one could safely exclude some false positives that kept appearing (i.e., the author name 'Paupy' (exclude -y) , the word depauperate (exclude -e), 'paupar-rash', 'paupal wing', 'Paupard' and 'Pauparding-Tritsch' (exclude -a) 'pAUPR' (exclude -r) 'Pauphalictus' (exclude -h)). Searches for ‘Winclada’ varied in their capitalization of the word but otherwise it was easy to find with 100% precision. 'Hennig86' was also referred to as 'Hennig 86' and that was easy to deal with. Some authors called TNT: 'TnT' or 'T.N.T.' thus these variations were incorporated into the search. Lots of other non-related concepts were unsurprisingly detected for 'TNT' so that search had low precision in the PLoS corpus. Finally NONA was also problematic for precision; some authors called it 'Nona' and this is harder to distinguish from the rice cultivar 'Nona Bokra' without special rules. Likewise, 'Nona' appeared as an author first name several times in PLoS papers, but I could not easily devise a safe general rule to exclude these false positives (Table 6.3).

Table 6.3: Summary of local desktop full-text searches. Most searches find references to intended software with 100% precision (no false positives).

Base Term	Corpus	Essence of Method	Total Hits (article)	Hits are intended concept (%)	Morph-based analyses
paup	BMC	grep -iR -l "paup[^ehyar]"	896	100	10
	PLOS		293	100	8
	Zootaxa		689	99.9	286
winclada	BMC	grep -iR -l "winclada"	7	100	1
	PLOS		3	100	0
	Zootaxa		151	100	96
hennig86	BMC	grep -iR -l "hennig86|hennig 86"	1	100	0
	PLOS		0	n/a	0
	Zootaxa		27	100	26
nona	BMC	grep -R -l "Nona[^A-Za-z]|NONA[^A-Za-z]"	10	90	3
	PLOS		12	16.6	0
	Zootaxa		103	96.1	88
tnt	BMC	grep -iR -l "tnt[^A-Za-z]|t\.n\.t"	7	57.1	2
	PLOS		190	7.9	5
	Zootaxa		151	100	105

phylip	BMC	grep -iR -l "phylip[^A-Za-z]"	912	100	1
	PLOS		241	100	0
	Zootaxa		22	100	6

If desired one can combine all the search terms with the boolean operators e.g. OR which in grep syntax is \| e.g. hennig86\|winclada. One can also print lines of context around the hit line to aid quick classification of the paper without having to open/read the original PDF. For example, with -A and -B in grep one can specify the number of lines of context to show above and below the line of interest. Samples of this output are given in Appendix 6.1.

Additional searches revealed that *just* looking in the reference list for references to the given software is not a safe, conservative strategy. For example:

find . -name "*.txt" -print0 | xargs -0 -n 1000 -P 3 grep -iR "goloboff" | grep -i "nixon"
as applied to the PLoS corpus found only 11 reference variants to “Goloboff PA, Farris JS, Nixon K (2003) TNT...”, and the year given varied from 2000-2008. The more thorough full-text search strategy for TNT given in Table6.3 gave many more false positives, but importantly it identified 15 instances in the PLoS corpus where TNT was used. Why the discrepancy? Manual examination of the four papers in which TNT was found but a citation including 'Goloboff' & 'Nixon' could not, demonstrated that:

- In Deo *et al* (2010) TNT was mis-cited as just “Goloboff PA (2000) TNT (Tree analysis using New Technology). 1.2 BETA ed. Tucuman, Argentina: By Authors.” Although primarily Goloboff's work, the official history of the program shows it has always been a collaborative work between Goloboff, Nixon & Farris. Also given the submission date of this paper (2009-11-24) I see no reason why (Goloboff *et al.* 2008) describing TNT was not cited.
- In Wilson (2010) TNT was misattributed to another paper by Goloboff: “Goloboff PA, Catalano SA, Mirande JM, Szumik CA, Arias JS, *et al.* (2009) Phylogenetic analysis of 73060 taxa corroborates major eukaryotic groups. Cladistics 25: 211–230”
- In Dilernia *et al* 2008 TNT was just referred to in the main text as “available at http://www.zmuc.dk/public/phylogeny/TNT)” which does not give proper attribution (WoK only counts citations given in the reference list).
- In Phillips *et al* (2010), TNT was misattributed to another paper by Goloboff: “Goloboff PA, Farris JS, Källersjö M, Oxelman B, Ramírez MJ, *et al.*. (2003)

Improvements to resampling measures of group support. Cladistics 19: 324–332”.

Table 6.4 Comparison of recall between popular academic web literature search portals, to the known number of articles that contain the search term in the full text of the article.

Journal	Period (inclusive)	Searched terms	MAS hits	MS hits	WoK hits	Scopus hits	GS hits	Local grep hits
Zootaxa	2001 to 2013-06-11	hennig86 OR hennig 86	0	0	0	13	10	25
Zootaxa	2001 to 2013-06-11	paup	0	2	6	444	332	705
Zootaxa	2001 to 2013-06-11	Nona OR NoName	0	2	5	75	49	90
Zootaxa	2001 to 2013-06-11	TNT OR T.N.T.	0	3	8	108	61	148
Zootaxa	2001 to 2013-06-11	phylip	0	0	0	14	8	21
Zootaxa	2001 to 2013-06-11	winclada	0	0	0	105	51	140
Zootaxa	2001 to 2013-06-11	phylogeny	0	292	1650	3903	2420	4596
Zootaxa	2001 to 2013-06-11	phylogen*	0	n/a	2104	5561	n/a	6804
Zootaxa	2001 to 2013-06-11	phylog*	0	n/a	2136	5618	n/a	6884
PLOS ONE	2006 to 2009	hennig86 OR hennig 86	0	0	0	0	0	0
PLOS ONE	2006 to 2009	paup	50	0	0	33	130	131
PLOS ONE	2006 to 2009	Nona OR NoName	1	0	0	1	8	6
PLOS ONE	2006 to 2009	TNT OR T.N.T.	10	2	2	7	81	84
PLOS ONE	2006 to 2009	phylip	58	1	1	20	99	100
PLOS ONE	2006 to 2009	winclada	0	0	0	0	2	2
PLOS ONE	2006 to 2009	phylogeny	257	507	521	774	680	678
PLOS ONE	2006 to 2009	phylogen*	9	2647	618	1087	n/a	1385
PLOS ONE	2006 to 2009	phylog*	0	2768	624	1105	n/a	1399

Table 6.4 gives the results of the recall tests for term searches between various academic search providers, compared against the known number of relevant article hits for that term in each journal (known from the local full-text searches, manually evaluated as part of work shown in Table 6.3). Example queries for each academic search provider are given in Appendix 6.2 for the purpose of transparency and reproducibility.

GS found most of the articles containing the searched-for terms in the open access journal PLoS One. In the case of the search for 'Nona' the discrepancy was caused by some false-positives returned: it finds three false-positive articles which have these words in them: “nona-L-arginine”, “nona-arginine” and “nonA”. My local full-text search for 'Nona/NONA/NoName' (Table 6.3) used a more sophisticated case-sensitive search to provide more precise results. This level of sophistry of search is not provided by any of the web-based academic search providers. The GS search results for Zootaxa, on average less than 44% recall, are significantly different in recall relevant to its performance against PLoS One, which had near 100% recall. Closer examination of the returned results for the GS Zootaxa searches show that the results returned often correlated very significantly with if the full-text of the paper had been deposited freely online at an institutional web address (see Appendix 6.3 for an example). MAS also showed a similar pattern of recall against PLoS One articles, presumably because MAS operates similarly to GS by crawling the web. By raw volume of returned results Scopus appears to be the best search provider for searching Zootaxa, but I would urge caution in this assessment because I did not check the precision of returned results in most cases, so these numbers of returned results could represent many false positives, the extent to which I do not know.

WoK unsurprisingly given it searches just titles-abstracts-keywords, for the specific software/methodological terms, retrieved hardly any articles but fared a little better for more general terms such as 'phylogeny', averaging just above 30% 'recall' if one generously assumes that none of the returned results were false positives. For the same general 'phylogeny' search terms Scopus had a much higher 'recall' averaging above 80%, again assuming no false positives.

6.5 Discussion

6.5.1 What exactly is WoK missing? Does it matter?

To determine the implications of WoK's apparent low recall against Zootaxa for general terms like phylogeny I sought to find and attempt to explain why it misses so many articles in which the word ‘phylogeny’ occured. To do this I manually examined the search results returned for the maximally-conservative search for 'phylog*' in Zootaxa by WoK for the

years 2005-2006 inclusive, which returned 172 articles (data in supplementary materials). I then scored if the term occurred in each of the articles in the abstract+title+keywords, main full text (excluding abs-title-key), and/or references.

Curiously, 45 of the WoK-found articles did not contain the search term phylog* in the abstract, title, or keywords (rather it was in the main text or references). Stranger still, of these 45, nine *only* had the searched-for term in the reference list at the end of the paper (Bray & Cribb 2005; Martin 2006; Samyn *et al.* 2006; Kajihara 2006; Sterrer 2006; Velez et al 2006; Dozsa-Farkas & Cech 2006; Winterton 2006; Craig et al 2006), which I would contend hardly makes the article of likely relevance to the initial search term.

I also checked the complement of articles that can be found to match 'phylog*' from a local desktop grep search, which returned 291 articles published in 2005, and 542 articles published in 2006. Of the 291 articles from 2005, 98 represented matches to a word or words in the reference-list only. This careful search of 291 articles published in 2005 revealed six articles (Li *et al.* 2005; Betancur-R & Acero 2005 ;Edgecombe & Hollington 2005; Sharkey 2005; Webb 2005; Ponder et al 2005) which contained new phylogenetic analyses reported in the paper that were not found by WoK searches for phylog*, phylogen* or phylogeny. I did not have time to manually search all 542 of the articles from 2006 that grep finds, but a cursory search also finds that WoK does not find Bravo (2006) because the language used in the abstract refers consistently to 'cladistics' rather than phylogenetics. An additional search for 'cladis*' in WoK only additionally finds Ponder et al (2005) and Bravo (2006), not the others.
This is highly significant as it demonstrates that a search strategy of relying on 'phylog*' OR 'cladis*' searches in WoK will not find at least 5 of the phylogenetic analyses (>10% of those published that year) that should be in the journal scope of WoK, let alone all the phylogenies ever published in all peer-reviewed journals.

6.5.2 What about the non-journal data archives? Do they have the data?

What of the phylogenetic data archives Treebase, Morphobank, and the more general data archive provided by Dryad? Could they be of help in finding phylogenetic data? Treebase

has data for 27 phylogenetic studies published in Zootaxa (http://treebase.org/treebase-web/search/studySearch.html?query=prism.publicationName==%22Zootaxa%22 data as accessed 2013-07-24). A simple *grep* of the Zootaxa corpus reveals 35 papers containing the string 'treebase' but interestingly these do not all match-up to the 27 in TreeBASE (data supplied in electronic supplementary materials: https://github.com/rossmounce/thesis_ESM/blob/master/lit_search_chapter/treebase_there_and_not.csv). One hit comes from the phrase 'treebased assessment'; nothing to do with TreeBASE. Another describes the TreeBASE project (Maddison *et al.* 2007). Another hit comes from Guayasamin *et al.* (2009) re-using published data in TreeBASE. In fact only 17 of those 35 'treebase' mentions in Zootaxa actually deposited their primary data in TreeBASE. There are seven studies (Triapitsyn *et al.* 2006; Harmer & Framenau, 2008; Kerr, 2010; Nygren *et al.* 2010; Brix *et al.* 2011; Prentice & Redak, 2012; Ballantyne & Lambkin, 2013;) which claim in the paper to have deposited data in TreeBASE, for which no corresponding public data can be found actually in TreeBASE. I do not know the reason behind this discrepancy but I have sometimes stumbled upon examples of this before with other papers in other journals during my PhD (e.g. Peach & Rouse (2004) which also claims to have associated deposited data in TreeBASE).

6.5.3 Few morphology-using phylogenetic analyses in BMC journals

The entire output of BMC published/owned journals from 2000-2011 contain just thirteen phylogenetic analyses derived (entirely or in part) from morphological data (Wagele & Kolb 2005; Asher 2007;Seiffert 2007;Haug *et al.* 2010; Zhao et al 2010;Struck 2007; Ahrens & Ribera 2009; Reidenbach et al 2009; Zrzavy et al 2009; Jenner et al 2009; Asher et al 2010; Pepato et al 2010; Geisler et al 2011) (the last eight of these are combined morphology and molecular analyses). The articles in BMC Evolutionary Biology & Frontiers in Zoology often required manual examination to determine. There are other studies that are closely related but not quite what I was seeking (e.g. numerous supertree studies, the phenetic UPGMA analysis of morphological measurements in Froufe *et al.* (2008), and Puniamoorthy *et al.* (2010) who generated a discrete character matrix but choose to map it onto pre-existing phylogenetic hypotheses). These studies all warranted close examination.

Applying the same approach on the BMC corpus did not find much, but was certainly worth doing: BMC publishes several general biology journals (e.g. Biology Direct, Journal of

Biology, BMC Research Notes, BMC Biology). It is conceivable that a relevant study could have been published in these journals, even if I subsequently found that not to be the case – the negative is worth establishing as much as the positive. The exercise also gave me a chance to discover interesting uses of cladistic methods in other disciplines e.g. Anthropology (Lycett, 2009).

Having done this analysis it would be remiss of me to ignore the bigger picture. Even with these full text methods, searching method sections written in natural language is harder than it needs to be. I would suggest that these phylogenetic method sections could easily be supplied as machine-readable statements using a controlled vocabulary of terms according to a reference standard (e.g. MIAPA [Leebens-Mack 2006]). The phylogenetic programs themselves could export the method used in an unambiguous, fully-explicit (no unmentioned hidden parameters) machine-readable format along with (or separate from), the data analysed with that method. In fact many phylogenetic programs already *can* export the data together with the results and the method, but for various reasons journals/authors/editors choose not to publish these in a re-usable, machine-readable form most of the time (Stoltzfus *et al.* 2012). Workflow tools like Armadillo (Lord *et al.* 2012) have also tried to enable reproducibility in phylogenetics by allowing researchers to save & record their entire workflow for prosperity and transparency but it remains seldom used so far (GS: 1 citation). The printed page should not and need not restrict what we do with scientific data. A multitude of high-quality options are available to authors who wish to make their data more easily re-usable including Dryad, Figshare, MorphoBank, Morphbank, and TreeBASE.

6.6 Conclusions

Whilst in other academic domains, rigorous full-text searches are the norm (e.g. medical sciences, c.f. The Cochrane Reviews), in ecology, systematics and evolution, rigorous approaches to literature search are seldom applied. This is partly because there is not adequate infrastructure (e.g., PubMed) to facilitate full-text full-corpus searching for the entire literature of this domain. Also it is prevented by legal restrictions and download rate limits that certain subscription access publishers place on 'their' material, making it very

hard to download all of it (Mounce, 2013). However, there is huge potential for use of full-text analysis in this area.

There are no longer technical barriers to large-scale data synthesis/mining in this area. I estimate, using Thomson Reuters JCR data, that the average journal in the areas of systematics, ecology and evolution publishes approximately 60 papers each year (median), and that a representative random sample of 200 journals from this subject area published cumulatively 17,000 articles last year [see Appendix 6.4]. If one scales this up this estimate to a 1,000 journals, from 2000-2013, using the mean PDF size given in 6.2 then the entire 'born-digital' literature for this area is likely to be well-less than 1.2 Terabytes (for perspective: commercially-available 3TB hard drives cost less than £80 as of 2013, with prices almost certain to decrease over time). This can of course be minimized further by just utilizing a HTML, XML or plain-text version rather than the PDF, but it shows one could easily archive all of the data on a standard desktop computer.

Chapter 7: Lost Branches in the Fossil Tree of Life

7.1 Introduction

In this concluding chapter I reflect-on and evaluate my experience of trying to find, re-extract, validate, and re-use published palaeomorphological data from the published literature. As I quickly found out during my research this is not an easy task and as such I feel a duty to report these findings in my thesis because I think my conclusions here are of real importance to future research efforts. To reconstruct some data sets from the published literature, the effort required is akin to that of fossil preparator delicately removing matrix from fossil bone – data is often difficultly 'embedded' in old scanned-in PDFs – one cannot just lift data out from these at the click of a button. Likewise, as fossil specimens are commonly found in a disarticulated and fragmentary state; the same is also true of palaeomorphological data in the literature. To 'save space' in the printed version of research publications much data is still commonly omitted, with only the newly added data being shown (and this is often replicated in the electronic version of the article). Yet to re-analyze or build on this data set one needs the *complete* data set. As I show in this chapter, this fragmented data availability is both unnecessary (the online version of the paper has no 'space' constraints other than file size, which the extra kilobytes of phylogenetic data should not trouble) and a significant hindrance. Emailing authors to ask them for a usable copy of their published data does not often result in a successful outcome. Drew *et al.* (2013) report that only 16% of 375 authors contacted, actually replied with the desired phylogenetic data. My own attempts at asking authors for their published data have had a similarly low success rate, although I did not care record the successes and failures in my inbox. However, I remain immensely grateful to those authors who *do* supply they published data upon request, or those who deposit their data online in an easily discoverable data repository like Dryad, TreeBASE or MorphoBank.

7.2 How many morphology-based phylogenetic studies are there?

The question I ask in this sub-heading is extremely important to establish as a baseline. Given that fossils and morphology are important to include in phylogenetic research (Chapter 1, Chapter 3) – how much of that type of research is there? Work that I contributed to; specifically the literature search and analysis of Stoltzfus *et al.* (2012) crudely demonstrates that around two thirds of papers that have the stemword “phylogen*” in them, as found by Web of Knowledge, report a phylogenetic analysis. Extrapolating this two thirds proportion to other publication years demonstrates that there are as a conservative estimate, over 100,000 published phylogenetic analyses in the modern 21st century peer-reviewed literature (2000-2012). Whilst the proportion of these that are morphology-based, or morphology-using (in conjunction with molecular data) is likely to be low, even then it should be well into the thousands. As I show in chapter 6, reliably discriminating between molecular and morphology-based phylogenetic analyses using currently available literature search techniques is extremely difficult. However, more sophisticated approaches using text and data mining techniques on full-text literature corpora may prove useful in this regard and it is an avenue of research I am actively pursuing thanks to the Panton Fellowship awarded to me by the Open Knowledge Foundation (Newman, 2012).

Throughout the last 4 years, I have been tagging wholly morphology-based vertebrate and invertebrate phylogenetic studies online, so that I have openly available bibliographic records of when and where they occur (Mounce 2013g,h). Reconciliation between the bibliographic data I hold, and that of Graeme Lloyd's bibliographic data on phylogenetic studies (Lloyd, 2009) shows that at a bare minimum, for the period 2000-2012 inclusive there are more than 3,800 peer-reviewed, published, morphology-based phylogenetic studies (at a bare minimum, I expect there are perhaps more than 5,000 if one also includes botanical studies which neither I nor Graeme Lloyd cover in much depth).

7.3 How much data from morphology-based phylogenetic studies is publicly available?

Given my estimate of 3,800 to >5,000 studies out there somewhere in the literature,

scattered across many hundreds of different journals (figure 6.1), a question that is easier to answer more definitively is: how many morphology-based phylogenetic studies are there in publicly available data repositories for immediate re-use? Thanks to William Piel (pers. comm.) there is a TreeBASE (Piel *et al.* 2002) query one can run that shows all morphology-using studies in TreeBASE:

http://treebase.org/treebase-web/search/matrixSearch.html?query=tb.type.matrix=Morphological%20or%20tb.type.matrix=Combination%20or%20tb.type.matrix=Behavior

Currently (2013-10-01) there are 784 matrices in TreeBASE found by this query, from 646 individual studies. 433 of the studies were published between 2000-2012. 635 of these matrices are solely morphological, 147 are combined data sets of morphology and molecules, and 2 are matrices of behavioural data. Note that for some, there's more than one matrix in TreeBASE corresponding to a single study.

Dryad (http://datadryad.org/) also archives phylogenetic data in a publicly available, re-usable format. It can pass-on data initially submitted to Dryad, to TreeBASE, so there is some duplication of content between the two. It is harder to search for purely morphological phylogenetic data sets in Dryad as it is a generic datastore for all sorts of biological data, but nevertheless I identify just 12 data sets in Dryad that are not yet in TreeBASE.

Morphobank (http://www.morphobank.org/ ; O'Leary and Kaufman, 2011) has 249 publicly available projects available as of 2013-10-01 but not all of these projects have an associated morphology-based phylogenetic data matrix. 218 of these represent studies from 2000-2012.

Thus between the big three publicly-accessible databases suitable for this type of data have morphological data from approximately 663 individual studies (assuming no duplication between Morphobank & TreeBASE content which is perhaps an unsafe assumption), for studies from the period 2000-2012. As I estimate there are comfortably 5000 such morphological studies in the literature from that time period – I estimate thus that we have programmatic access to data from just 13% of the studies from this recent period.

But what of data provided in the article or the supplementary materials? This data is often error-prone (e.g. typesetting errors), ill-formatted, inextractable and difficult to re-use. Nor is it very discoverable (Chapter 6). It is difficult to explain on paper quite how much of a dumping ground supplementary materials files are, so I instead point to my Young Systematists' Forum talk on this very matter for further evidence (Mounce, 2010).

7.4 Replicating published cladistic analyses

Interestingly, I have observed many times during my work that the published dataset, and the published methodology of analysis e.g. parameters, character orderings and weightings, do NOT match the published results. Most of the time the differences are small, but for more than just a few papers the difference really is striking and significantly contradicts at least some of the conclusions of a paper.

One such example that I successfully challenged was that of the phylogeny published in a high-profile, front cover Nature article by Liu *et al.* (2011). The data simply didn't match the reported result no matter what parameters one applies, this was further confirmed by another research group who also independently noticed the analysis was not reproducible (Legg *et al.* 2011). I include my published article (Mounce & Wills, 2011) in full, in Appendix 7.1 to further evidence this.

7.5 Comments on the Liu *et al.* reply

The authors of the original article wrote a formal reply (Liu *et al.* 2011b) to both our re-analysis (Mounce & Wills 2011) and that of Legg *et al.* (2011). In this interesting reply they accept many of our criticisms but do not appear to readily admit the main point of Mounce & Wills (2011) and Legg *et al.* (2011) that the consensus tree supported by their data is significantly different from the one they initially reported. Figure 1 (Liu *et al.* 2011b) presents what they purport to be a bootstrap analysis of their data to “verify the stability” of their findings. Bootstrap analyses (Felsenstein 1985) resample characters with replacement. Given their data matrix only has 38 characters most of which are binary, for 28 taxa – it's clear *a priori* that bootstrap support values are extremely unlikely to be 100 for each and every node of the consensus tree. Alas, I report here that much like their

previous analyses (Liu *et al.* 2011a) I could not replicate their bootstrap analysis in either relationships depicted or strength of support for the optimal topology. I can only conclude the people who contributed to the data analysis of both Liu *et al.* (2011a,b) have a very different interpretation of cladistics and bootstrap analyses to most other people in the research in community. Similarly they called the PTP test they used the 'partitioning tail permutation' test, yet myself and most others know it as the *permutation tail probability* test (Faith & Cranston 1991).

7.5 Suggestions for the future of data publication and review

Research data is immensely valuable, including re-use value. Yet current publishing practices don't seem to me to be treating it as such – scarcely any data is made available for re-use as this chapter demonstrates. Most of the focus behind the publication and review process appears to go on the paper and the figures, not the code, data and analyses behind them. I have suggested we should change this (e.g. at the Systematics Association Biennial Meeting; Mounce, 2011b). In morphology-based phylogenetics where the analyses are typically computationally very simple – just seconds to find the consensus tree with TNT for most parsimony analyses – that reviewers should routinely re-run the data to independently check and validate the reported results in such papers. This would hopefully go some way to prevent problems of non-reproducibility like that reported in Mounce & Wills (2011). Furthermore, if the authors hand-over the raw re-usable data at review for the reviewers to access, it would seem relatively simple to me to ensure that once the article is accepted, that this raw data file also gets made publicly available in an appropriate data repository e.g. TreeBASE, Dryad or Morphobank. This would be a huge improvement on the current jumble of supplementary information files that are rarely immediately re-usable – it is a far cry from the simplicity and utility of Genbank for DNA sequence data. Usually supplementary files are PDF's mixing lots of different types of data in a long document, or a Word file, or even a spreadsheet. By putting data in a data repository it ensures scalable programmatic access and the chance to build rich searchable metadata on top of the data deposit (e.g. as TreeBASE does) to make the data more easily discoverable. Data buried on the 90th page of a supplementary file PDF is effectively lost to re-use to all but the most determined of data re-users.

7.6 Positive reasons to share published data

I would like to end on a positive note. The scholarly publishing landscape is rapidly changing at the moment. The US, UK, Europe and countries around the world are now committed to moving towards open access publishing. Many journals have adopted a *mandatory* data archiving policy for all authors e.g *Evolution* (Fairbairn 2011). Research funders are explicitly recognising the value of data and code, not just publications (Piwowar, 2013). Research is demonstrating that data archiving is a worthwhile investment (Piwowar *et al.* 2011) and that sharing detailed research data is associated with an increased citation rate (Piwowar *et al.*, 2007). The latest results show that papers with open data available with no legal restrictions on its usage receive approximately 9% more citations than similar studies in which data are not made available (Piwowar & Vision 2013). I don't doubt that this apparent citation benefit is transferable to palaeontology, systematics, macroevolutionary studies and ecology – data re-use is a common facet of science. Instead of being mandated or 'forced' to share data, perhaps one day researchers will be eager to share data, to their own benefit as well as that of others (Poisot *et al.* 2013). With publications such as White *et al.* (2013) providing clear advice and help on making data re-usable, I'm confident that in future we'll have less 'lost branches' (sensu Drew *et al.* 2013) on the Fossil Tree of Life.

References

Aberer, A. J., Krompass, D., and Stamatakis, A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology* 62:162-166.

Adnet, S. and Cappetta, H. 2001. A palaeontological and phylogenetical analysis of squaliform sharks (Chondrichthyes: Squaliformes) based on dental characters. *Lethaia* 34:234-248.

Adrain, J. M. 2001. Systematic Paleontology. *Journal of Paleontology* 75:1055-1057.

Ahyong, S.T. 2006. Phylogeny of the clawed lobsters (Crustacea: Decapoda: Homarida) *Zootaxa* 1109:1-14

Allain, R. and Aquesbi, N. 2008. Anatomy and phylogenetic relationships of *Tazoudasaurus naimi* (Dinosauria, Sauropoda) from the late Early Jurassic of Morocco. *Geodiversitas* 30:345–424.

Allard, M.W., Farris, J.S. and Carpenter, J.M. 1999a. Congruence among mammalian mitochondrial genes. *Cladistics* 15:75–84.

Allard, M.W., Kearney, M., Kivimaki, K.L., Meisner, A.D. and Strong, E.E. 1999b. The random cladist: A review of the software package Random Cladistics. *Cladistics* 15:183–189.

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Thomas, Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N., and Bunce, M. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences* 279:4724-4733.

Anderson, J.S., Reisz, R.R., Scott, D., Frobisch, N.B. and Sumida, S.S. 2008. A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders. *Nature* 453:515–518.

Andres, B., Clark, J. M., and Xing, X. 2010. A new rhamphorhynchid pterosaur from the Upper Jurassic of Xinjiang, China, and the phylogenetic relationships of basal pterosaurs. *Journal of Vertebrate Paleontology* 30:163-187.

Apesteguia, S. and Zaher, H. 2006. A Cretaceous terrestrial snake with robust hindlimbs and a sacrum. *Nature* 440:1037-1040.

Arango, C. P. and Wheeler, W. C. 2007. Phylogeny of the sea spiders (Arthropoda, Pycnogonida) based on direct optimization of six loci and morphology. *Cladistics* 23:255-293

Archie, J. W. 1989. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Systematic Biology* 38:253–269.

Archie, J.W. 1996. Measures of Homoplasy pp. 153-188. In: M.J. Sanderson, L. Hufford, editors. Homoplasy: The Recurrence of Similarity in Evolution. Academic Press, New York.

Archie, J.W. and Felsenstein J. 1993. The number of evolutionary steps on random and

minimum length trees for random evolutionary data. *Theoretical Population Biology* 43:52–79.

Arratia, G. 2009. Identifying patterns of diversity of the actinopterygian fulcra. *Acta Zoologica* 90:220–235.

Asher, R.J., Meng, J., Wible, J.R., McKenna, M.C., Rougier, G.W., Dashzeveg, D. and Novacek, M.J. 2005. Stem Lagomorpha and the antiquity of Glires. *Science* 307:1091–1094.

Asher, R.J. and Hofreiter, M. 2006. Tenrec phylogeny and the noninvasive extraction of nuclear DNA. *Systematic Biology* 55:181–194.

Asher, R.J. 2007. A web-database of mammalian morphology and a reanalysis of placental phylogeny. *BMC Evolutionary Biology* 7:108+.

Asher, R.J., Bennett, N. and Lehmann, T. 2009. The new framework for understanding placental mammal evolution. *Bioessays* 31:853-864.

Ashlock, P. D. 1974. The uses of cladistics. *Annual Review of Ecology and Systematics* 5: 81-99.

Austin, J. J., Ross, A. J., Smith, A. B., Fortey, R. A., and Thomas, R. H. 1997. Problems of reproducibility – does geologically ancient DNA survive in amber–preserved insects? *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264:467-474.

Avise, J.C. and Robinson, T.J. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Systematic Biology* 57:503–507.

Babot, J., D. A. López, and T. J. Gaudin 2012. The most ancient xenarthran petrosal: morphology and evolutionary significance. *Journal of Vertebrate Paleontology* 32:1186-1197.

Ballantyne, L. A. and Lambkin, C. L. 2013. Systematics and Phylogenetics of Indo-Pacific Luciolinae Fireflies (Coleoptera: Lampyridae) and the Description of new Genera. *Zootaxa* 3653:1-162.

Baker, R., Yu, X. and DeSalle, R. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Molecular Phylogenetics and Evolution* 9:427–436.

Bakker, M. and Wicherts, J. M. 2011. The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods* 43:666-678.

Barker, F.K. and Lutzoni, F.M. 2002. The utility of the incongruence length difference test. *Systematic Biology* 51:625–37.

Baum, B.R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41:3–10.

Beard, K. C., Marivaux, L., Chaimanee, Y. , Jaeger, J.-J., Marandat, B., Tafforeau, P., Soe, A.N., Tun, S.T. And Kyaw, A.A. 2009. A new primate from the Eocene Pondaung formation of Myanmar and the monophyly of Burmese amphipithecids. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 276:3285–3294.

Beck, R. M. D., Godthelp, H., Weisbecker, V. Archer, M. and Hand, S.J. 2008. Australia's oldest marsupial fossils and their biogeographical implications. *PLoS ONE* 3:e1858+.

Beckmann, M. and von Wehrden, H. 2012. Where you search is what you get: literature

mining – google scholar versus web of science using a data set from a literature search in vegetation science. *Journal of Vegetation Science* 23:1197-1199.

Bennett, S. C. 2013. The phylogenetic position of the Pterosauria within the Archosauromorpha re-examined. *Historical Biology* 25:545-563.

Bergsten, J. 2005 A review of long-branch attraction. *Cladistics* 21:163-193.

Berle, D. and Starcevic V. 2007 Inconsistencies between reported test statistics and p-values in two psychiatry journals. *International Journal of Methods in Psychiatric Research* 16: 202-207.

Betancur-R, R. and Acero, A. P. 2005. Description of Cathorops mapale, a new species of sea catfish (Siluriformes: Ariidae) from the Colombian Caribbean, based on morphological and mitochondrial evidence. *Zootaxa* 1045:45-60.

Beutel, R. G., S.-q. Ge, and T. Hörnschemeyer 2008. On the head morphology of Tetraphalerus, the phylogeny of Archostemata and the basal branching events in Coleoptera. *Cladistics* 24 (3), 270-298.

Beutel, R. G., B. Wang, J.-J. Tan, S.-Q. Ge, D. Ren, and X.-K. Yang 2012. On the phylogeny and evolution of Mesozoic and extant lineages of Adephaga (Coleoptera, Insecta). *Cladistics* 29:147-165.

Bisconti, M. 2008. Morphology and phylogenetic relationships of a new eschrichtiid genus (Cetacea: Mysticeti) from the Early Pliocene of northern Italy. *Zoo. J. Linn. Soc.* 153:161-186.

Blagoderov, V. and D. Grimaldi 2004. Fossil Sciaroidea (Diptera) in Cretaceous ambers, exclusive of Cecidomyiidae, Sciaridae, and Keroplatidae. *American Museum Novitates*, 1-76.

Blagoderov, V., Hippa, H., and Sevcik, J. 2009. Asiorrhina, a new oriental genus of Lygistorrhinidae (Diptera: Sciaroidea) and its phylogenetic position. *Zootaxa* 2295:31-45.

Bledsoe, A. H. and Raikow, R. J. 1990. A quantitative assessment of congruence between molecular and nonmolecular estimates of phylogeny. *Journal of Molecular Evolution* 30:247-259.

Bloch, J.I., Silcox, M.T. Boyer, D.M. and Sargis, E.J. 2007. New paleocene skeletons and the relationship of plesiadapiforms to crown-clade primates. *Proceedings of the National Academy of Sciences of the United States of America* 104:1159–1164.

Bloech, C., Weiss-Schneeweiss, H., Schneeweiss, G. M., Barfuss, M. H., Rebernig, C. A., Villaseñor, J. L. L., and Stuessy, T. F. 2009. Molecular phylogenetic analyses of nuclear and plastid DNA sequences support dysploid and polyploid chromosome number changes and reticulate evolution in the diversification of melampodium (Millerieae, Asteraceae). *Molecular Phylogenetics and Evolution* 53:220-233.

Bluis, J. and Shin, D.G. 2003. Nodal distance algorithm: Calculating a phylogenetic tree comparison metric. In: BIBE '03: Proceedings of the 3rd IEEE Symposium on BioInformatics and BioEngineering. Washington, DC, USA: IEEE Computer Society. p. 87+

Boessenecker, R. W. and Churchill, M. 2013. A reevaluation of the morphology, paleoecology, and phylogenetic relationships of the enigmatic walrus Pelagiarctos. *PLoS ONE* 8:e54311+.

Bonde, N. 1976. Cladistic classification as applied to Vertebrates pp. 741–804. In: Hecht, M.K., Goody, P.C. and Hecht, B.M. (eds.) Major Patterns in Vertebrate Evolution. Plenum Press, New York.

Bouetel, V. and de Muizon, C. 2006. The anatomy and relationships of Piscobalaena nana (Cetacea, Mysticeti), a Cetotheriidae s.s. from the Early Pliocene of Peru. *Geodiversitas* 28:319-395.

Bourdon, E., Ricqles, A. and Cubo, J. 2009. A new transantarctic relationship: morphological evidence for a Rheidae–Dromaiidae–Casuariidae clade (Aves, Palaeognathae, Ratitae). *Zoological Journal of the Linnean Society* 156:641–662.

Bourque, M. 1978. Arbres de steiner et reseaux dont certains sommets sont a localisation variable. Thesis, University of Montreal.

Bravo, F. 2006. The taxonomy of Neotropical Brunettiina (Diptera, Psychodidae, Psychodinae, Mormiini), with descriptions of ten new species from Brazil and comments on the generic classification of the subtribe. *Zootaxa* 1134:1-28.

Bremer, B. 1996. Combined and separate analyses of morphological and molecular data in the plant family Rubiaceae. *Cladistics* 12:21–40.

Brix, S., Riehl, T. and Leese, F. 2011. First genetic data for species of the genus Haploniscus Richardson, 1908 (Isopoda: Asellota: Haploniscidae) from neighbouring deep-sea basins in the South Atlantic. *Zootaxa* 2838:79-84.

Brochu, C. A., Njau, J., Blumenschine, R. J., and Densmore, L. D. 2010. A new horned crocodile from the Plio-Pleistocene hominid sites at olduvai gorge, Tanzania. *PLoS ONE* 5:e9333+.

Brundin, L. 1966. Transantarctic relationships and their significance, as evidenced by chironomid midges. *K. Svenska Vet. Akad. Handl.* 11:1-472.

Brusatte, S. L., Benson, R. B. J., and Xu, X. 2012. A reassessment of *Kelmayisaurus petrolicus*, a large theropod dinosaur from the Early Cretaceous of China. *Acta Palaeontologica Polonica* 57:65-72.

Bryant, H. N. 1995. Why autapomorphies should be removed: A reply to Yeates. *Cladistics* 11:381–384.

Buckley, T. R., Simon, C., Shimodaira, H. and Chambers, G. K. 2001. Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (Maoricicada) using multiple-comparison tests of tree topology. *Molecular Biology and Evolution* 18:223-234.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L. and Waddell, P.J. 1993. Partitioning and combining data in phylogenetic analysis. *Systematic Biology* 42:384–397.

Butler, R.J., Upchurch, P. and Norman, D.B. 2008. The phylogeny of the Ornithischian dinosaurs. *Journal of Systematic Palaeontology* 6:1–40.

California Digital Library 2013. SciVerse Hub is Retiring in August 2013. http://www.cdlib.org/cdlinfo/2013/06/17/sciverse-hub-is-retiring-in-august-2013/ [Webpage]

Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A., and Poinar, G. O. 1993. Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* 363:536-538.

Carballido, J. L., Garrido, A. C., Canudo, J. I., and Salgado, L. 2010. Redescription of *Rayososaurus agrioensis* Bonaparte (Sauropoda, Diplodocoidea), a rebbachisaurid from the early Late Cretaceous of Neuquén. *Geobios* 43:493-502.

Carbone, I., Anderson, J. B., and Kohn, L. M. (1999). Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. *Evolution* 16: 354 –362.

Carpenter, J.M. 1988. Choosing among multiple equally parsimonious cladograms. *Cladistics* 4:291–296.

Carrano, M.T. and Sampson, S.D. 2008. The phylogeny of Ceratosauria (Dinosauria: Theropoda). *Journal of Systematic Palaeontology* 6:183–236.

Carrasco, P. A., Mattoni, C. I., Leynaud, G. C., and Scrocchi, G. J. 2012. Morphology, phylogeny and taxonomy of South American bothropoid pitvipers (serpentes, viperidae). *Zoologica Scripta* 41:109-124.

Carroll, H., Ebbert, M., Clement, M. & Snell, Q. 2007. PSODA: Better tasting and less filling than PAUP. Proceedings of the 4th Biotechnology and Bioinformatics Symposium, pp 74-78

Carroll, H., Teichert, A. R., Krein, J., Sundberg, K., Snell, Q., and Clement, M. 2009. An open source phylogenetic search and alignment package. *International journal of bioinformatics research and applications* 5:349-364.

Cau, A. and Fanti, F. 2011. The oldest known metriorhynchid crocodylian from the middle Jurassic of north-eastern Italy: *Neptunidraco ammoniticus* gen. et sp. nov. *Gondwana Research* 19: 550-565.

Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., and Ebrahim, N. A. 2013. A comparison between two main academic literature collections: Web of Knowledge and Scopus databases. *Asian Social Science* 9.

Cheng, L., Chen, X., Zeng, X., and Cai, Y. 2012. A new Eosauropterygian (Diapsida: Sauropterygia) from the Middle Triassic of Luoping, Yunnan province. *Journal of Earth Science* 23:33-40.

Claeson, K. M., O'Leary, M. A., Roberts, E. M., Sissoko, F., Bouaré, M., Tapanila, L., Goodwin, D., and Gottfried, M. D. 2010. First mesozoic record of the stingray Myliobatis wurnoensis from Mali and a phylogenetic analysis of Myliobatidae incorporating dental characters. *Acta Palaeontologica Polonica* 55:655-674.

Clarke, J. A., D. T. Ksepka, N. A. Smith, and M. A. Norell 2009. Combined phylogenetic analysis of a new north american fossil species confirms widespread Eocene distribution for stem rollers (Aves, Coracii). *Zoological Journal of the Linnean Society* 157 (3), 586-611.

Clarke, D. J. 2011. Testing the phylogenetic utility of morphological character systems, with a revision of *Creophilus* Leach (Coleoptera: Staphylinidae). *Zoological Journal of the Linnean Society* 163:723-812.

Clennett, J. C. B., Chase, M. W., Forest, F., Maurin, O., and Wilkin, P. 2012. Phylogenetic systematics of erythronium (liliaceae): morphological and molecular analyses. *Botanical Journal of Linnean Society* 170:504-528.

Cobbett A., Wilkinson, M. and Wills, M.A. 2007. Fossils impact as hard as living taxa in parsimony analyses of morphology. *Systematic Biology* 56:753-766.

Collard, M., Gibbs, S. and Wood, B. 2001. Phylogenetic utility of higher primate postcranial morphology. *American Journal of Physical Anthropology* Supplement 32: 52.

Colless, D.H. 1980. Congruence between morphometric and allozyme data for Menidia species: A reappraisal. *Systematic Zoology* 29:288-299.

Conrad, J. L. 2008. Phylogeny and systematics of Squamata (Reptilia) based on morphology. *Bulletin of the American Museum of Natural History* 310:1-182.

Conway Morris, S. 1977. A new metazoan from the Cambrian Burgess Shale of British Columbia. *Palaeontology* 20:623-640.

Csiki, Z., Vremir, M., Brusatte, S. L., and Norell, M. A. 2010. An aberrant island-dwelling theropod dinosaur from the Late Cretaceous of Romania. *Proceedings of the National Academy of Sciences* 107:15357-15361.

Cunningham, C. W. 1997a Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* 14(7):733-740.

Cunningham, C. W. 1997b Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Systematic Biology* 46: 464-478.

D'Erchia, A.M., Gissi, C., Pesole, G., Saccone, C.and Arnasonn, U. 1996. The guinea-pig is not a rodent. *Nature* 381:597–600.

Davis, M. P. 2010. Evolutionary relationships of the Aulopiformes (Euteleostei: Cyclosquamata): a molecular and total evidence approach pp. 431-470. In: J.S. Nelson, H.P. Schultze and M.V.H. Wilson (eds.) Origin and Phylogenetic Interrelationships of Teleosts. Verlag Dr. Friedrich Pfeil, Munchen, Germany.

Daza, J. D., Alifanov, V. R., and Bauer, A. M. 2012. A redescription and phylogenetic reinterpretation of the fossil lizard *Hoburogekko suchanovi* Alifanov, 1989 (Squamata, Gekkota), from the Early Cretaceous of Mongolia. *Journal of Vertebrate Paleontology* 32:1303-1312.

De Pietri, V. L., Costeur, L., Güntert, M., and Mayr, G. 2011. A revision of the Lari (Aves, Charadriiformes) from the Early Miocene of Saint-Gérand-le-Puy (Allier, France). *Journal of Vertebrate Paleontology* 31:812-828.

de Vienne, D.M., Giraud, T. and Martin, O.C. 2007. A congruence index for testing topological similarity between trees. *Bioinformatics* 23:3119–3124.

Degnan, J.H. and Rosenberg, N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24:332–340.

Deng, T. 2008. A new elasmothere (Perissodactyla, Rhinocerotidae) from the Late Miocene of the Linxia basin in Gansu, China. *Geobios* 41:719-728.

Deo, A. J., Costa, R., DeLisi, L. E., DeSalle, R., and Haghighi, F. 2010. A novel analytical framework for dissecting the genetic architecture of behavioral symptoms in neuropsychiatric disorders. *PLoS ONE* 5:e9714+.

Dietze, K. 2009. Morphology and phylogenetic relationships of certain neoteleostean fishes from the Upper Cretaceous of Sendenhorst, Germany. *Cretaceous Research* 30:559-574.

Dilernia, D. A., Jones, L., Rodriguez, S., Turk, G., Rubio, A. E., Pampuro, S., Gomez-Carrillo, M., Bautista, C., Deluchi, G., Benetucci, J., Lasala, M. B., Lourtau, L., Losso,

M. H., Perez, H., Cahn, P., and Salomón, H. 2008. HLA-driven convergence of HIV-1 viral subtypes b and f toward the adaptation to immune responses in human populations. *PLoS ONE* 3:e3429+.

Diogo, R. 2004. Morphological Evolution, Adaptations, Homoplasies, Constraints, and Evolutionary Trends: Catfishes as a Case Study on General Phylogeny & Macroevolution. Science Publishers, Enfield.

Diogo, R. 2007. On the origin and evolution of higher-clades: osteology, myology, phylogeny and macroevolution of bony fishes and the rise of tetrapods. Enfield, Science Publishers.

Diogo, R., Doadrio, I., and Vandewalle, P. 2008. Teleostean phylogeny based on osteological and myological characters. *International Journal of Morphology* 26:463-522.

Dockes, J-F. 2013. Recoll. http://www.lesbonscomptes.com/recoll/ [accessed 2013-06-25]

Dolphin, K., Belshaw, R., Orme, C.D.L. and Quicke, D.L.J. 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. *Molecular Phylogenetics and Evolution* 17:401–406.

Dong, X.-P. P., Donoghue, P. C., Cunningham, J. A., Liu, J.-B. B., and Cheng, H. 2005. The anatomy, affinity, and phylogenetic significance of *Markuelia*. *Evolution & Development* 7:468-482.

Donoghue, M. J., Doyle, J. A., Gauthier, J., Kluge, A. G., and Rowe, T. 1989. The importance of fossils in phylogeny reconstruction. *Annual Review of Ecology and Systematics* 20:431-460.

Donoghue, M.J. and Ree, R.H. 2000. Homoplasy and developmental constraint: A model and an example from plants. *American Zoologist* 40:759-769.

Donoghue, P. C. J. and Benton, M. J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology and Evolution* 22:424–431.

Doolittle, W.F. and Bapteste, E. 2007. Pattern pluralism and the tree of life hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 104:2043–2049.

Dowton, M. and Austin, A. D. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy-the behavior of the incongruence length difference test in mixed-model analyses. *Systematic Biology* 51:19-31.

Draper, I., Hedenas, L. and Grimm, G. L. 2007. Molecular and morphological incongruence in European species of Isothecium (Bryophyta). *Molecular Phylogenetics and Evolution* 42:700–716.

Drew, B. T., Gazis, R., Cabezas, P., Swithers, K. S., Deng, J., Rodriguez, R., Katz, L. A., Crandall, K. A., Hibbett, D. S., and Soltis, D. E. 2013. Lost branches on the tree of life. *PLoS Biology* 11:e1001636+.

Eddy, D. R. and Clarke, J. A. 2011. New information on the cranial anatomy of *Acrocanthosaurus atokensis* and its implications for the phylogeny of Allosauroidea (Dinosauria: Theropoda). *PLoS ONE* 6:e17932+. [PMCID: PMC3061882]

Edgecombe, G. D. and Hollington, L. M. 2005. Morphology and relationships of a new species of *Henicops* (Chilopoda: Lithobiomorpha) from New South Wales and Queensland, Australia. *Zootaxa* 961:1-20.

Edgecombe, G. D. 2010. Palaeomorphology: fossils and the inference of cladistic relationships. *Acta Zoologica* 91:72–80.

Eernisse, D. J. and Kluge, A. G. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Molecular Biology and Evolution* 10:1170-1195.

Egan, M.G. 2006 Support versus corroboration. *Journal of Biomedical Informatics* 39:72-85.

Eichholz, 1949. Aristotle's Theory of the Formation of Metals and Minerals . The Classical Quarterly 43:141-146.

Erwin, D. H. 2008. Extinction as the loss of evolutionary history. *Proceedings of the National Academy of Sciences of the United States of America* 105 Suppl 1:11520-11527.

Estabrook, G.F. 1968. A general solution in partial orders for the Camin-Sokal model in phylogeny. *Journal of Theoretical Biology* 21:421–438.

Estabrook, G.F., McMorris, F.R. and Meacham, C.A. 1985. Comparison of undirected trees based on subtrees of four evolutionary units. *Systematic Zoology* 34:193–200.

Ezcurra, M.D. and Cuny, G. 2007. The coelophysoid *Lophostropheus airelensis*, gen. nov.: A review of the systematics of *"Liliensternus" airelensis* from the Triassic-Jurassic outcrops of Normandy (France). *Journal of Vertebrate Paleontology* 27:73+.

Fairbairn, D. J. 2011. The advent of mandatory data archiving. *Evolution* 65:1-2.

Faith, D.P. and Cranston, P.S. 1991. Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. *Cladistics* 7:1–28.

Faivovich, J. 2002. A cladistic analysis of Scinax (Anura: Hylidae). *Cladistics* 18:367–393.

Farke, A. A., Ryan, M. J., Barrett, P. M., Tanke, D. H., Braman, D. R., Loewen, M. A., and Graham, M. R. 2011. A new centrosaurine from the late cretaceous of alberta, canada, and the evolution of parietal ornamentation in horned dinosaurs. *Acta Palaeontologica Polonica* 56:691-702.

Farris, J.S. 1971. The hypothesis of nonspecificity and taxonomic congruence. *Annual Review of Ecology and Systematics* 2:277–302.

Farris, J.S. 1973. On comparing the shapes of taxonomic trees. *Systematic Zoology* 22:50–54.

Farris, J. S. 1976. Phylogenetic classification of fossils with recent species. *Systematic Zoology* 25:271-282.

Farris, J. S. 1989a. The retention index and the rescaled consistency index. *Cladistics* 5: 417-419.

Farris, J.S. 1989b. The retention index and homoplasy excess. *Systematic Biology* 38:406–407.

Farris, J.S. 1991. Excess Homoplasy Ratios. *Cladistics* 7: 81-91.

Farris, J. S. 1997. Combinability and congruence. *Cladistics* 13, 170. [Conference Abstract]

Farris, J.S., Kluge, A.G. and Eckardt, M.J. 1970. A numerical approach to phylogenetic systematics. *Systematic Zoology* 19:172-189.

Farris, J.S., Källersjö, M., Kluge, A.G. and Bult, C. 1995a. Constructing a significance test for incongruence. *Systematic Biology* 44:570–572.

Farris, J.S., Källersjö, M., Kluge, A.G. and Bult, C. 1995b. Testing significance of incongruence. *Cladistics* 10:315–319.

Felsenstein, L. 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 39:783-791.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22:521–565.

Fidler, F., Burgman, M. A., Cumming, G., Buttrose, R., and Thomason, N. 2006. Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology. *Conservation Biology* 20:1539-1544.

Fikáček, M., Prokin, A., Angus, R. B., Ponomarenko, A., Yue, Y., Ren, D., and Prokop, J. 2012. Phylogeny and the fossil record of the helophoridae reveal jurassic origin of extant hydrophiloid lineages (Coleoptera: Polyphaga). *Systematic Entomology* 37:420-447.

Finarelli, J. 2008. A total evidence phylogeny of the Arctoidea (Carnivora: Mammalia): Relationships among basal taxa. *Journal of Mammalian Evolution* 15:231-259.

Finarelli, J.A. and Clyde, W.C. 2004. Reassessing hominoid phylogeny: Evaluating congruence in the morphological and temporal data. *Paleobiology* 30: 614-651.

Fingerman, S. 2004. Elsevier Holds Official Launch Events for Scopus. http://newsbreaks.infotoday.com/NewsBreaks/Elsevier-Holds-Official-Launch-Events-for-Scopus-16330.asp

Fisher-Reid, C. and Wiens, J. 2011. What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? lessons from plethodon salamanders and 13 other vertebrate clades. *BMC Evolutionary Biology* 11:300+.

Fitzgerald, E.M.G. 2010. The morphology and systematics of Mammalodon colliveri (Cetacea: Mysticeti), a toothed mysticete from the Oligocene of Australia. *Zoological Journal of the Linnean Society* 158: 367-476.

Ford, L. and O'Hara, L. H. 2008. It's all academic: Google scholar, scirus, and windows live academic search. *Journal of Library Administration* 46:43-52.

Fordyce, R. E. and F. G. Marx 2013. The pygmy right whale *Caperea marginata*: the last of the cetotheres. *Proceedings of the Royal Society B: Biological Sciences* 280 (1753)

Forey, P. L. and Fortey, R. A. 2001. Fossils in the reconstruction of phylogeny. Pp. 515–519 in D. E. G. Briggs & P. R. Crowther (editors), Palaeobiology II. Blackwell Science, Inc., Malden, Massachusetts.

Forey, P. L. 2004. Systematics and Paleontology. In: D. M. Williams and P. L. Forey (eds.) Milestones in Systematics. Systematics Association Special Volume, CRC Press, 149-180.

Forey, P.L. 2005. Cladistics For Palaeontologists. Palaeontology Newsletter http://www.palass.org/modules.php?name=palaeo&page=19&sec=newsletter

Frazzetta, T. H. 2012. Flatfishes, turtles, and bolyerine snakes: Evolution by small steps or large, or both? *Evolutionary Biology* 39: 30-60.

Friedman, M. 2007. Styloichthys as the oldest coelacanth: Implications for early

osteichthyan interrelationships. *Journal of Systematic Palaeontology* 5:289–343.

Friedman, M. 2008. The evolutionary origin of flatfish asymmetry. *Nature* 454:209–212.

Fritz, U., Branch, W. R., Hofmeyr, M. D., Maran, J., Prokop, H., Schleicher, A., Iroký, Stuckas, H., Vargas-Ramírez, M., Vences, M., and Hundsdörfer, A. K. 2011. Molecular phylogeny of african hinged and helmeted terrapins (testudines: Pelomedusidae: Pelusios and pelomedusa). *Zoologica Scripta* 40:115-125.

Frobisch, J. 2007. The cranial anatomy of Kombuisia frerensis Hotton (Synapsida, Dicynodontia) and a new phylogeny of anomodont therapsids. *Zoological Journal of the Linnean Society* 150:117–144.

Fu, J. and Murphy, R.W. 1999 Discriminating and locating character covariance: an application of permutation tail probability (PTP) analyses. *Systematic Biology* 48: 380-395.

Gaffney, E.S., Hooks, G.E. and Schneider, V.P. 2009. New material of North American side-necked turtles (Pleurodira: Bothremydidae). *American Museum Novitates* 3655:1–26.

Gaffney, E. S. and Krause, D. W. 2011. Sokatra, a new Side-Necked turtle (Late Cretaceous, Madagascar) and the diversification of the main groups of pelomedusoides. *American Museum Novitates* 1-28.

Garcia-Perez, M. 2012. Reviewer's report of: Gehanno et al (2013) Is Google Scholar sensitive enough to be used alone for systematic reviews? http://www.biomedcentral.com/imedia/8023217788144832_comment.pdf

Garwood, R. J., Dunlop, J. A., Giribet, G., and Sutton, M. D. 2011. Anatomically modern carboniferous harvestmen demonstrate early cladogenesis and stasis in opiliones. *Nature Communications* 2:444+.

Gates, T.A. and Sampson, S.D. 2007. A new species of *Gryposaurus* (Dinosauria: Hadrosauridae) from the late Campanian Kaiparowits formation, Southern Utah, USA. *Zoological Journal of the Linnean Society* 151:351–376.

Gaubert, P., Wozencraft, W.C., Cordeiro-Estrela, P. and Veron, G. 2005. Mosaics of convergences and noise in morphological phylogenies: What's in a viverrid-like carnivoran? *Systematic Biology* 54:865–894.

Gaudin, T., Emry, R. and Wible, J. 2009. The phylogeny of living and extinct pangolins (Mammalia, Pholidota) and associated taxa: A morphology based analysis. *Journal of Mammalian Evolution* 16:235–305.

Gauthier, J., Kluge, A. G., and Rowe, T. 1988. Amniote phylogeny and the importance of fossils. *Cladistics* 4:105-209.

Gehanno, J. F., Rollin, L., and Darmoni, S. 2013. Is the coverage of google scholar enough to be used alone for systematic reviews? *BMC Medical Informatics and Decision Making* 13:7+.

Gilbert, C. C. 2013. Cladistic analysis of extant and fossil african papionins using craniodental data. *Journal of Human Evolution* 64:399-433.

Giribet, G. 2010. A new dimension in combining data? the use of morphology and phylogenomic data in metazoan systematics. *Acta Zoologica* 91:11-19.

Giustini, D. 2013. Is Google scholar enough for SR searching? No.

http://blogs.ubc.ca/dean/2013/01/is-google-scholar-enough-for-sr-searching-no/ [Blog post]

Goddard, W., Kubicka, E., Kubicki, G. and McMorris, F.R. 1994. The agreement metric for labelled binary trees. *Mathematical Biosciences* 123:215–226.

Godefroit, P., Shulin, H., Tingxiang, Y. and Lauters, P. 2008. New hadrosaurid dinosaurs from the uppermost Cretaceous of Northeastern China. *Acta Palaeontologica Polonica* 53:47–74.

Goldman, N., Anderson, J.P. and Rodrigo, A.G. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49:652–670.

Goloboff P. A. 1999a. Analyzing large data sets in reasonable times: Solutions for composite optima. *Cladistics* 15:415-428.

Goloboff, P.A. 1999b. NONA (NO NAME) ver. 2 Published by the author, Tucumán, Argentina.

Goloboff, P. A., Farris, J. S., and Nixon, K. C. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24:774-786.

Goloboff, P.A. 2008. Calculating SPR distances between trees. *Cladistics* 24:591–597.

Gonzalez-Riga, B., Previtera, E. and Pirrone, C. 2009. *Malarguesaurus florenciae* gen. et sp. nov., a new titanosauriform (Dinosauria, Sauropoda) from the Upper Cretaceous of Mendoza, Argentina. *Cretaceous Research* 30:135–148.

Google 2013. Our history in depth. http://www.google.com/about/company/history/

Gould, G.C. 2001. The phylogenetic resolving power of discrete dental morphology among extant hedgehogs and the implications for their fossil record. *American Museum Novitates* 3340:1-52.

Graham, D. 2013 Academic Publishing: Survey of funders supports the benign Open Access outcome priced into shares. *HSBC Global Research* 1-36

Grant, T., and Kluge, A.G. 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics* 19: 379-418.

Grantham, T. 2004. The role of fossils in phylogeny reconstruction: Why is it so difficult to integrate paleobiological and neontological evolutionary biology? *Biology and Philosophy* 19:687-720.

Greenwood, D. R., Hill, C. R., and Conran, J. G. 2013. *Prumnopitys anglica* sp. nov. (Podocarpaceae) from the Eocene of England. *Taxon* 62(3) 565-580.

Guang-Hui, X. U. and Mee-Mann Chan, G. 2009. Redescription of *Paralycoptera wui* Chang & Chou, 1977 (Teleostei: Osteoglossoidei) from the Early Cretaceous of eastern China. *Zoo. J. of the Linn. Soc.* 157:83-106.

Guayasamin, J. M., Castroviejo-Fisher, S., Trueb, L., Ayarzaguena, J., Rada, M., Vila, C. Phylogenetic systematics of Glassfrogs (Amphibia: Centrolenidae) and their sister taxon *Allophryne ruthveni*. *Zootaxa* 2100:1-97.

Gutiérrez, G. and Marín, A. 1998. The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Molecular Biology and Evolution* 15:926-929.

Harmer, A. M. T. and Framenau, V. W. 2008. *Telaprocera*, a new Australian orb-web spider

genus with elongated webs (Araneae: Araneidae). *Zootaxa* 1956, 59-80.

Hastings, A.K., Bloch, J.I. and Jaramillo, C.A. 2011. A new longirostrine dyrosaurid (Crocodylomorpha, Mesoeucrocodylia) from the Paleocene of north-eastern Colombia: Biogeographic and behavioural implications for New-World Dyrosauridae. *Palaeontology* 54:1095-1116.

Hebsgaard, M. B., Phillips, M. J., and Willerslev, E. 2005. Geologically ancient DNA: fact or artefact? *Trends in Microbiology* 13:212-220.

Hedges, S.B. and Sibley, C.G. 1994. Molecules vs. morphology in avian evolution: the case of the “pelecaniform” birds. *Proceedings of the National Academy of Sciences of the United States of America* 91:9861–9865.

Hedges, S.B. and Maxson, L.R. 1996. Re: Molecules and morphology in amniote phylogeny. *Molecular Phylogenetics and Evolution* 6:312-314.

Hennig, W. 1950. Grundzüge einer Theorie der phylogenetischen Systematik, Berlin: Deutscher Zentralverlag

Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.

Hill, R.V. 2005. Integration of morphological data sets for phylogenetic analysis of Amniota: The importance of integumentary characters and increased taxonomic sampling. *Systematic Biology* 54:530–547.

Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* 383:140–141.

Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47:3–8.

Hilton, E. J. and P.L. Forey. 2009. Redescription of ?*Chondrosteus acipenseroides* Egerton, 1858 (Acipenseriformes, ?Chondrosteidae) from the lower Lias of Lyme Regis (Dorset, England), with comments on the early evolution of sturgeons and paddlefishes. *Journal of Systematic Palaeontology* 7:427–453.

Hinchliff, C. E. and Roalson, E. H. 2013. Using supermatrices for phylogenetic inquiry: an example using the sedges. *Systematic Biology* 62:205-219.

Hipp, A.L., Hall, J.C. and Sytsma, K.J. 2004. Congruence versus phylogenetic accuracy: Revisiting the incongruence length difference test. *Systematic Biology* 53:81–89.

Holland, T. and Long, J.A. 2009. On the phylogenetic position of *Gogonasus andrewsae*, within the Tetrapodomorpha. *Acta Zoologica* 90: 285-296.

Hoot, S. B., Magallon, S., and Crane, P. R. (1999). Phylogeny of basal eudicots based on three molecular data sets: atpB, rbcL and 18S nuclear ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 86: 1–32.

Hospitaleche, C.A., Tambussi, C., Donato, M. and Cozzuol, M. 2007. A new Miocene penguin from Patagonia and its phylogenetic relationships. *Acta Palaeontologica Polonica* 52:299–314.

Hoyal Cuthill, J. F., Braddy, S. J., and Donoghue, P. C. J. 2010. A formula for maximum possible steps in multistate characters: isolating matrix parameter effects on measures of evolutionary convergence. *Cladistics* 26:98-102.

Huelsenbeck, J. P. 1991. When are fossils better than extant taxa in phylogenetic analysis? *Systematic Zoology* 40:458-469.

Hull, D.L. 1990. Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. University of Chicago Press, Chicago.

Hunt, G. 2008. Gradual or pulsed evolution: When should punctuational explanations be preferred? *Paleobiology* 34: 360-377.

Hurley, I.A., Mueller, R.L., Dunn, K.A., Schmidt, E.J., Friedman, M., Ho, R.K., Prince, V.E., Yang, Z., Thomas, M.G. and Coates, M.I. 2007. A new time-scale for ray-finned fish evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 274:489–498.

Huson, D.H. and Scornavacca, C. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* 61:1061-1067.

Hutchinson, M. N., Skinner, A., and Lee, M. S. Y. 2012. *Tikiguania* and the antiquity of squamate reptiles (lizards and snakes). *Biology Letters* 8:665-669.

Imamura, H., Shirai, S.M. and Yabe, M. 2005. Phylogenetic position of the family Trichodontidae (Teleostei: Perciformes), with a revised classification of the perciform suborder Cottoidei. *Ichthyological Research* 52:264–274.

Janssens, S. B., Wilson, Y. S., Yuan, Y.-M., Nagels, A., Smets, E. F., and Huysmans, S. 2012. A total evidence approach using palynological characters to infer the complex evolutionary history of the asian *Impatiens* (Balsaminaceae). *Taxon* 61:355-367.

Jenner, R. A., Dhubhghaill, C. N., Ferla, M. P., and Wills, M. A. 2009. Eumalacostracan phylogeny and total evidence: limitations of the usual suspects. *BMC Evolutionary Biology* 9:21+.

Jenner, R. A. 2004a. Accepting partnership by submission? morphological phylogenetics in a molecular millennium. *Systematic Biology* 53:333-359.

Jenner, R. A. 2004b. When molecules and morphology clash: reconciling conflicting phylogenies of the metazoa by considering secondary character loss. *Evolution & Development* 6:372-378+.

Ji, Q., Z.-X. Luo and Ji, S.-A. 1999. A Chinese triconodon mammal and mosaic evolution of the mammalian skeleton. *Nature* 398: 326-330.

Jinha, A. E. 2010. Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing* 23:258-263.

Johnson, K. P. and Sorensen, M. D. 1998. Comparing molecular evolution in two mitochondrial protein coding genes (cytochrome b and ND2) in the dabbling ducks (Tribe: Anatini). *Molecular Phylogenetics and Evolution* 10: 82-94.

Jouve, S., Iarochene, M., Bouya, B. and Amaghzaz, M. 2006. A new species of *Dyrosaurus* (Crocodylomorpha, Dyrosauridae) from the early Eocene of Morocco: Phylogenetic implications. *Zoological Journal of the Linnean Society* 148:603–656.

Joyce, W.G. 2007. A phylogeny of Mesozoic turtles. *Bull. of the Peabody Museum of Natural History* 48:3-102

Joyce, W. G. and Sterli, J. 2012. Congruence, non-homology, and the phylogeny of basal turtles. *Acta Zoologica* 93:149-159.

Karasawa, H., Schweitzer, C.E., 2006. A new classification of the Xanthoidea sensu lato (Crustacea: Decapoda: Brachyura) based on phylogenetic analysis and traditional systematics and evaluation of all fossil Xanthoidea sensu lato. *Contr. Zool.* 75:23–73.

Karasawa, H., Schweitzer, C.E., Feldmann, R.M. 2011. Phylogenetic analysis and revised classification of podotrematous Brachyura (Decapoda) including extinct and extant families. *J. Crust. Biol.* 31:523-565.

Kempthorne, O. 1952. Design and Analysis of Experiments. Wiley, New York.

Kerr, P. H. (2010) Phylogeny and classification of Rhagionidae, with implications for

Tabanomorpha (Diptera: Brachycera). *Zootaxa* 2592:1–133.

Kishino, H. and Hasegawa, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* 29:170-179.

Kitching, I., Forey, P., Humphries, C. and Williams, D. 1998. Cladistics: Theory and Practice of Parsimony Analysis (Systematics Association Special Volume). Oxford University Press, Oxford.

Klassen, G. J., Mooi, R.D. and Locke, A. 1991. Consistency indices and random data. *Systematic Zoology* 40: 446-457.

Klingenberg, C.P. 2008. Morphological integration and developmental modularity. *Annual Reviews of Ecology, Evolution and Systematics* 39:115-132.

Klug, S. 2010. Monophyly, phylogeny and systematic position of the Synechodontiformes (Chondrichthyes, Neoselachii). *Zoologica Scripta* 39:37-49.

Kluge, A. G. and Farris, J. S. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18:1-32.

Kluge, A.G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). *Systematic Zoology* 38:7–25.

Klymiuk, A. A. and Stockey, R. A. 2012. A Lower Cretaceous (Valanginian) seed cone provides the earliest fossil record for *Picea* (Pinaceae). *American Journal of Botany* 99:1069-1082.

Koperwas J., Walczak K. 2011. Tree edit distance for leaf-labelled trees on free leafset and its comparison with frequent subsplit dissimilarity and popular distance measures. *BMC Bioinformatics* 12:204+.

Koski, K. 2007. The mandibular complex. *European Journal of Orthodontics* 29: i118-i123.

Kroh, A. and Smith, A. B. 2010. The phylogeny and classification of post-Palaeozoic echinoids. *Journal of Systematic Palaeontology* 8:147-212.

Ksepka, D.T. 2009. Broken gears in the avian molecular clock: New phylogenetic analyses support stem galliform status for *Gallinuloides wyomingensis* and rallid affinities for *Amitabha urbsinterdictensis*. *Cladistics* 25:173–197.

Ksepka, D. T. and Clarke, J. A. 2010. New fossil mousebird (Aves: Coliiformes) with feather preservation provides insight into the ecological diversity of an Eocene north american avifauna. *Zoological Journal of the Linnean Society* 160:685-706.

Ksepka, D. T., J. A. Clarke, S. J. Nesbitt, F. B. Kulp, and L. Grande 2013. Fossil evidence of wing shape in a stem relative of swifts and hummingbirds (aves, Pan-Apodiformes). *Proceedings of the Royal Society B: Biological Sciences* 280:20130580+.

Kumar, S., Skjaeveland, A., Orr, R., Enger, P., Ruden, T., Mevik, B.H., Burki, F., Botnen, A. and Tabrizi, K.S. 2009. AIR: A batch-oriented web program package for construction of

supermatrices ready for phylogenomic analyses. *BMC Bioinformatics* 10:357+.

Kurochkin, E. N. 1996. A new enantiornithid of the Mongolian Late Cretaceous, and a general appraisal of the Infraclass Enantiornithes (Aves). *Russian Academy of Sciences* 60pp.

Lambert, O. 2008. A new porpoise (Cetacea, Odontoceti, Phocoenidae) from the Pliocene of the North Sea. *Journal of Vertebrate Paleontology* 28:863-872.

Lambert, O., de Muizon, C., and Bianucci, G. 2013. The most basal beaked whale *Ninoziphius platyrostris* Muizon, 1983: clues on the evolutionary history of the family Ziphiidae (Cetacea: Odontoceti). *Zool J Linn Soc* 167:569-598.

Langille, M. G. I. and Eisen, J. A. 2010. BioTorrents: A file sharing service for scientific data. *PLoS ONE* 5:e10071+

Le Quesne, W.J. 1969. A method of selection of characters in numerical taxonomy. *Systematic Biology* 18:201–205.

Lee, M.S.Y. 2001a. Molecules, morphology, and the monophyly of diapsid reptiles. *Contributions to Zoology* 70:1-22.

Lee, M.S.Y. 2001b. Uninformative characters and apparent conflict between molecules and morphology. *Molecular Biology and Evolution* 18:676–680.

Lee, M.S.Y. and Scanlon, J. D. 2002. Snake phylogeny based on osteology, soft anatomy and ecology. *Biological Reviews* 77:333–401.

Leebens-Mack, J., Vision, T., Brenner, E., Bowers, J. E., Cannon, S., Clement, M. J., Cunningham, C. W., dePamphilis, C., deSalle, R., Doyle, J. J., Eisen, J. A., Gu, X., Harshman, J., Jansen, R. K., Kellogg, E. A., Koonin, E. V., Mishler, B. D., Philippe, H., Pires, J. C., Qiu, Y.-L. L., Rhee, S. Y., Sjölander, K., Soltis, D. E., Soltis, P. S., Stevenson, D. W., Wall, K., Warnow, T., and Zmasek, C. 2006. Taking the first steps towards a standard for reporting on phylogenies: Minimum information about a phylogenetic analysis (MIAPA). *Omics : a journal of integrative biology* 10:231-237.

Legg, D. A., Ma, X., Wolfe, J. M., Ortega-Hernandez, J., Edgecombe, G. D., and Sutton, M. D. 2011. Lobopodian phylogeny reanalysed. *Nature* 476:E1.

Legg, D. A., Sutton, M. D., Edgecombe, G. D., and Caron, J.-B. 2012. Cambrian bivalved arthropod reveals origin of arthrodization. *Proceedings of the Royal Society B: Biological Sciences* 279:4699-4704.

Lehtonen, S., Sääksjärvi, I. E., Ruokolainen, K., and Tuomisto, H. 2011. Who is the closest extant cousin of humans? total-evidence approach to hominid phylogenetics via simultaneous optimization. *Journal of Biogeography* 38:805-808.

Leigh, J.W., Lapointe, F.J., Lopez, P. and Bapteste, E. 2011. Evaluating phylogenetic congruence in the post-genomic era. *Genome Biology and Evolution* 3:571-587.

Leighton, L.R. and Maples, C.G. 2002. Evaluating internal versus external characters: Phylogenetic analyses of the Echinoconchidae, Buxtoniinae, and Juresaniinae (Phylum Brachiopoda). *Journal of Paleontology* 76:659–671.

Li, C., Lee, J.S. Groebner, J.L *et al.* 2005.

Li, P.-P., Gao, K.-Q., Hou, L.-H. and Xu, X. 2007. A gliding lizard from the early Cretaceous of China. *Proceedings of the National Academy of Sciences of the United States of America* 104:5507–5509.

Lin, Y., Rajan, V. and Moret, B. 2011. A metric for phylogenetic trees based on matching bioinformatics research and applications. *Lecture Notes in Computer Science* 6674:197-208

Linnaeus, C. 1745. Gothländska Resa på Riksens Högloflige Ständers Befallningf förrättad åhr 1741. *Stockholm & Uppsala*. [in Swedish]

Linnaeus, C. 1758. Systema Naturae, Ed. 10, Vol. 1. 824 pp. Salvii, Holmiae.

Lister, A. M., Edwards, C.J., Nock, D.A.W., Bunce, M., van Pijlen, I.A., Bradley, D.G., Thomas, M.G. and Barnes, I. 2005. The phylogenetic position of the 'giant deer' *Megaloceros giganteus*. *Nature* 438:850–853.

Liu, J., Steiner, M., Dunlop, J. A., Keupp, H., Shu, D., Ou, Q., Han, J., Zhang, Z., and Zhang, X. 2011a. An armoured cambrian lobopodian from China with arthropod-like appendages. *Nature* 470:526-530.

Liu, J., Steiner, M., Dunlop, J. A., Keupp, H., Shu, D., Ou, Q., Han, J., Zhang, Z., and Zhang, X. 2011b. Liu et al. reply. *Nature* 476:E1.

Lloyd, G. T. 2009. Matrices http://www.graemetlloyd.com/matr.html

Lopardo, L., Giribet, G., and Hormiga, G. 2011. Morphology to the rescue: molecular data and the signal of morphological characters in combined phylogenetic analyses—a case study from mysmenid spiders (Araneae, Mysmenidae), with comments on the evolution of web architecture. *Cladistics* 27:278-330.

Lopez, P. and Bapteste, E. 2009. Molecular phylogeny: reconstructing the forest. *Comptes Rendus Biologies* 332:171-182.

López-Antoñanzas, R., Flynn, L. J., and Knoll, F. 2013. A comprehensive phylogeny of extinct and extant rhizomyinae (Rodentia): evidence for multiple intercontinental dispersals. *Cladistics* 29:247-273.

Lord, E., Leclercq, M., Boc, A., Diallo, A. B., and Makarenkov, V. 2012. Armadillo 1.1: An original workflow platform for designing and conducting phylogenetic analysis and simulations. *PLoS ONE* 7:e29903+.

Lu, J., Unwin, D.M., Jin, X., Liu, Y. and Ji, Q. 2010. Evidence for modular evolution in a long-tailed pterosaur with a pterodactyloid skull. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 277:383–389.

Luo, Z.-X. X., Yuan, C.-X. X., Meng, Q.-J. J., and Ji, Q. 2011. A jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* 476:442-445.

Lutzoni, F.M. 1997. Phylogeny of lichen- and non-lichen-forming omphalinoid mushrooms and the utility of testing for combinability among multiple data sets. *Systematic Biology* 46:373–406.

Lyson, T.R. and Joyce, W.G. 2009. A new species of *Palatobaena* (Testudines: Baenidae) and a maximum parsimony and Bayesian phylogenetic analysis of Baenidae. *Journal of Paleontology* 83:457–470.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46:523–536.

Maddison, D. R., Schulz, K-S., and Maddison, W. P. The Tree of Life Web Project. *Zootaxa* 1668:19-40.

Manegold, A. and Töpfer, T. 2012. The systematic position of hemicircus and the stepwise evolution of adaptations for drilling, tapping and climbing up in true woodpeckers

(picinae, picidae). *Journal of Zoological Systematics and Evolutionary Research*

Manos, P. S., Soltis, P. S., Soltis, D. E., Manchester, S. R., Oh, S.-H., Bell, C. D., Dilcher, D. L., and Stone, D. E. 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Systematic Biology* 56:412-430.

Mariadassou, M., Bar-Hen, A., and Kishino, H. 2012. Taxon influence index: assessing taxon-induced incongruities in phylogenetic inference. *Systematic Biology* 61:337-345.

Martinelli, A.G. and Rougier, G.W. 2007. On *Chaliminia musteloides* (Eucynodontia: Tritheledontidae) from the Late Triassic of Argentina, and a phylogeny of Ictidosauria. *Journal of Vertebrate Paleontology* 27:442–460.

Martinez, R.N. and Alcober, O.A. 2009. A basal sauropodomorph (Dinosauria: Saurischia) from the Ischigualasto formation (Triassic, Carnian) and the early evolution of Sauropodomorpha. *PLoS ONE* 4:e4397+.

Matsumoto, R., Suzuki, S., Tsogtbaatar, K. and Evans, S. 2009. New material of the enigmatic reptile *Khurendukhosaurus* (Diapsida: Choristodera) from Mongolia. *Naturwissenschaften* 96:233–242.

Mayor, A. 2000. The First Fossil Hunters Paleontology in Greek and Roman Times. Princeton University Press.

Mayr, E. 1953. Methods and Principles of Systematic Zoology. McGraw Hill, New York.

Mayr, G. 2005. A new cypselomorph bird from the Middle Eocene of Germany and the early diversification of avian aerial insectivores. *The Condor* 107:342-352.

Mayr, G., Rana, R. S., Rose, K. D., Sahni, A., Kumar, K., Singh, L., and Smith, T. 2010. *Quercypsitta*-Like birds from the Early Eocene of India (Aves, ?psittaciformes). *Journal of Vertebrate Paleontology* 30:467-478.

Mayr, G. 2010b. Phylogenetic relationships of the paraphyletic 'caprimulgiform' birds (nightjars and allies). *Journal of Zoological Systematics and Evolutionary Research* 48:126-137.

Mayr, G. 2011a. The phylogeny of charadriiform birds (shorebirds and allies) – reassessing the conflict between morphology and molecules. *Zoological Journal of the Linnean Society* 161:916–934.

Mayr, G. 2011b. Cenozoic mystery birds – on the phylogenetic affinities of bony-toothed birds (Pelagornithidae). *Zoologica Scripta* 40:448-467.

Mayr, G. and Mourer-Chauviré, C. 2004. Unusual tarsometatarsus of a mousebird from the Paleogene of France and the relationships of *Selmes* Peters, 1999. *Journal of Vertebrate Paleontology* 24: 366-372.

Maurício, G. N., Areta, J. I., Bornschein, M. R., and Reis, R. E. 2012. Morphology-based phylogenetic analysis and classification of the family Rhinocryptidae (Aves: Passeriformes). *Zoological Journal of the Linnean Society* 166:377-432.

McDonald, A. T., Kirkland, J. I., DeBlieux, D. D., Madsen, S. K., Cavin, J., Milner, A. R. C., and Panzarin, L. 2010. New basal iguanodonts from the cedar mountain formation of utah and the evolution of Thumb-Spiked dinosaurs. *PLoS ONE* 5:e14075+.

Meier, R., Kores, P., and Darwin, S. 1991. Homoplasy slope ratio: A better measurement of observed homoplasy in cladistic analyses. *Systematic Zoology* 40.

Mendeley 2013. Compare Mendeley. http://www.mendeley.com/compare-mendeley/

Michener, C.D. 1953. Life-history studies in insect systematics. *Systematic Zoology* 2:112-118.

Mickevich, M.F. 1978. Taxonomic congruence. *Systematic Zoology* 27:143–158.

Mickevich, M.F. and Farris J.S. 1981. The implications of congruence in Menidia. *Systematic Zoology* 30:351-370.

Microsoft 2013. About Microsoft Academic Search http://academic.research.microsoft.com/About/help.htm#5

Mihaljevic M., Jerjen I. and A.B. Smith 2011. The test architecture of *Clypeaster* (Echinoidea, Clypeasteroida) and its phylogenetic significance. *Zootaxa* 2983: 21-38

Mitteroecker, P. and Bookstein, F. 2007. The conceptual and statistical relationship between modularity and morphological integration. *Systematic Biology* 56:818-836.

Mitteroecker, P. and Bookstein, F. 2008. The evolutionary role of modularity and integration in the hominoid cranium. *Evolution* 62:943-958.

Miyamoto, M. M. and Fitch, W. M. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* 44:64-76.

Mooi, R.D. and Gill, A.C. 2010. Phylogenies without synapomorphies — a crisis in fish systematics: Time to show some character. *Zootaxa* 2450:26–40.

Moore, M. J., Hassan, N., Gitzendanner, M. A., Bruenn, R. A., Croley, M., Vandeventer, A., Horn, J. W., Dhingra, A., Brockington, S. F., Latvis, M., Ramdial, J., Alexandre, R., Piedrahita, A., Xi, Z., Davis, C. C., Soltis, P. S., and Soltis, D. E. 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: Insights into Deeper-Level angiosperm relationships from a long, slowly evolving sequence region. *International Journal of Plant Sciences* 172:541-558.

Mounce, R. 2010. The continued growth of phylogenetic information. Young Systematists' Forum, 2010, London http://prezi.com/1s0lkatmc30t/the-continued-growth-of-phylogenetic-information/ [Conference Slides]

Mounce, R. 2011a. Phylogenetic congruence between cranial and postcranial characters in archosaur systematics. Society of Vertebrate Paleontology Annual Meeting, p163. [Conference Abstract]

Mounce, R. 2011b. Nullius in Calculo. Talk presented at the Eleventh Systematics Association Biennial Meeting, Dublin. http://prezi.com/rwyvfb9bvec_/nullius-in-calculo/

Mounce, R. [data] 2013a. BMC phylogeny paper corpus. https://github.com/rossmounce/BMC_phylogeny_corpus

Mounce, R. [data] 2013b. Article PDF file sizes. https://github.com/rossmounce/thesis_ESM/blob/master/lit_search_chapter/forR_ArticlepdfFileSizes.csv

Mounce, R. [data] 2013c. Web of Knowledge Winclada search data https://github.com/rossmounce/thesis_ESM/blob/master/lit_search_chapter/wok_winclada.bib

Mounce, R. [data] 2013d. Scopus Winclada search data https://github.com/rossmounce/thesis_ESM/blob/master/lit_search_chapter/scopus_winclada.bib

Mounce, R. [data], 2013e. Phylogeny publishers data exported from Web of Knowledge https://github.com/rossmounce/thesis_ESM/blob/master/lit_search_chapter/phylogen_publishers_in_wok.csv

Mounce, R. [data], 2013f. Longtail data on the distibution of phylogeny articles https://github.com/rossmounce/thesis_ESM/blob/master/lit_search_chapter/long_tail_phylogen_articles_data.csv

Mounce, R. [data] 2013g. Bibliographic data for vertebrate morphology-based phylogenetic studies. http://www.citeulike.org/user/rossmounce/tag/morphophylogeny

Mounce, R. [data] 2013h. Bibliographic data for invertebrate morphology-based phylogenetic studies. http://www.citeulike.org/user/rossmounce/tag/nvmorphophylogeny

Mounce, R. 2013. Content Mining. Talk given in Brussels at the Licences for Europe Stakeholder dialogue for the European Commission. http://www.slideshare.net/rossmounce/content-mining [Conference Talk]

Muller, J. and Reisz, R.R. 2006. The phylogeny of early eureptiles: Comparing parsimony and Bayesian approaches in the investigation of a basal fossil clade. *Systematic Biology* 55:503–511.

Naylor, G.J.P. and Adams, D.C. 2001. Are the fossil data really at odds with the molecular data? Morphological evidence for Cetartiodactyla phylogeny reexamined. *Systematic Biology* 50: 444-453.

Near, T. J., Pesavento, J.J. and Cheng, C.-H.C. 2003. Mitochondrial DNA, morphology, and the phylogenetic relationships of antarctic icefishes (Notothenioidei: Channichthyidae). *Molecular Phylogenetics and Evolution* 28: 87-98.

Near, T.J. 2009. Conflict and resolution between phylogenies inferred from molecular and phenotypic data sets for hagfish, lampreys, and gnathostomes. *Journal of Experimental Zoology* 312:749–761.

Nee, S., and May, R. M. 1997. Extinction and the loss of evolutionary history. *Science* 278:692:694.

Nelson, G.J., “Cladism” as a philosophy of classification. *Systematic Zoology* 20:373–376, 1971.

Nelson, G.J. 1969. Origin and diversification of teleostean fishes *Annals of the New York Academy of Sciences* 167: 18–30.

Nelson, G.J. 1970. Outline of a theory of comparative biology, Systematic Zoology 19:373–384.

Newman, L. 2012. Introducing our Panton Fellows. http://blog.okfn.org/2012/03/30/introducing-our-panton-fellows/

Ngamskulrungroj, P., Gilgado, F., Faganello, J., Litvintseva, A. P., Leal, A. L., Tsui, K. M., Mitchell, T. G., Vainstein, M. H., and Meyer, W. 2009. Genetic diversity of the cryptococcus species complex suggests that *Cryptococcus gattii* deserves to have varieties. *PLoS ONE* 4:e5862+.

Nichols, R. 2001. Gene trees and species trees are not the same. *Trends in Ecology and Evolution* 16:358–364.

Nixon K. C. 1999. The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics* 15:407-414.

Nixon, K. C. 2002. WinClada, version 1.00. 08. Published by the author, Ithaca, New York.

Novacek, M. J. 1992. Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. *Systematic Biology* 41:58-73.

Novacek, M. J., and Wheeler, Q. D. 1992. Extinct taxa—accounting for 99.9990 . . . % of the earth's biota. pp. 1–16 in Extinction and Phylogeny. M. J. Novacek and Q. D. Wheeler (eds.) Columbia University Press, New York.

Nygren, A., Sundkvist, T., Mikac, B. and Pleijel, F. 2010. Two new and two poorly known autolytines (Polychaeta:Syllidae) from Madeira and the Mediterranean Sea. *Zootaxa* 2640:35-52.

O'Connor, P.M., Sertich, J.J.W., Stevens, N.J., Roberts, E.M., Gottfried, M.D., Hieronymus, T.L., Jinnah, Z.A., Ridgely, R., Ngasala, S.E. and Temba, J. 2010. The evolution of mammal-like crocodyliforms in the Cretaceous period of Gondwana. *Nature* 466: 748-751.

O'Leary, M.A., Gatesy, J., and Novacek, M.J. 2003. Are the dental data really at odds with the molecular data? Morphological evidence for whale phylogeny (Re)Reexamined. *Systematic Biology* 52:853–864.

O'Leary, M. A. and Kaufman, S. 2011. MorphoBank: phylophenomics in the " cloud". Cladistics 27:529-537.

O'Leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., Goldberg, S. L., Kraatz, B. P., Luo, Z.-X. X., Meng, J., Ni, X., Novacek, M. J., Perini, F. A., Randall, Z. S., Rougier, G. W., Sargis, E. J., Silcox, M. T., Simmons, N. B., Spaulding, M., Velazco, P. M., Weksler, M., Wible, J. R., and Cirranello, A. L. 2013. The placental mammal ancestor and the post-K-pg radiation of placentals. *Science* 339:662-667.

O'Meara, B. 2009. TreeBASE mirror http://www.brianomeara.info/treebase [accessed December 12th 2009]

Olson, E. 1991. George Gaylord Simpson 1902-1984, A Biographical Memoir. National Academy of Sciences.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A. M. V., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G. D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J. F., Weinstock, J., Gregersen, K., Roed, K. H., Eisenmann, V., Rubin, C. J., Miller, D. C., Antczak, D. F., Bertelsen, M. F., Brunak, S., Al-Rasheid, K. A. S., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, Kjaer, K., Sicheritz-Ponten, T., Jensen, L. J., Olsen, J. V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., and Willerslev, E. 2013. Recalibrating equus evolution using the genome sequence of an early middle pleistocene horse. *Nature* 499:74-78.

Osi, A. and Makadi, L. 2009. New remains of *Hungarosaurus tormai* (Ankylosauria, Dinosauria) from the Upper Cretaceous of Hungary: Skeletal reconstruction and body mass estimation. *Paleontologische Zeitschrift* 83:227–245.

Page, R.D.M. 1996. On consensus, confidence, and 'total evidence'. *Cladistics* 12:83–92.

Pamilo, P., Nei, M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5:568–583.

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in r language. Bioinformatics 20:289-290.

Parenti, L.R. 2008. A phylogenetic analysis and taxonomic revision of ricefishes, oryzias and relatives (Beloniformes, Adrianichthyidae). *Zoological Journal of the Linnean Society* 154:494–610.

Pattengale, N.D., Gottlieb, E.J. and Moret, B.M.E. 2007. Efficiently computing the Robinson-Foulds metric. *Journal of Computational Biology* 14:724–735.

Patterson, C. 1981. Significance of fossils in determining evolutionary relationships. *Annual Review of Ecology and Systematics* 12:195-223.

Peach, M.B. and Rouse G.W. 2004. Phylogenetic trends in the abundance and distribution of pit organs of elasmobranchs. *Acta Zoologica* 85:233-244

Peñalver, E., J. Ortega-Blanco, A. Nel, and X. Delclòs 2010. Mesozoic evaniidae (Insecta: Hymenoptera) in spanish amber: Reanalysis of the phylogeny of the evanioidea. *Acta Geologica Sinica* 84:809-827.

Pepato, A., da Rocha, C., and Dunlop, J. 2010. Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evolutionary Biology* 10:235+.

Perrichot, V., Nel, A., and Quicke, D. 2009. New braconid wasps from french Cretaceous amber (Hymenoptera, Braconidae): synonymization with eoichneumonidae and implications for the phylogeny of ichneumonoidea. *Zoologica Scripta* 38:79-88.

Phillips, M.J., Bennett, T.H. and Lee, M.S.Y. 2009. Molecules, morphology, and ecology indicate a recent, amphibious ancestry for echidnas. *Proceedings of the National Academy of Sciences of the United States of America* 106:17089–17094.

Phillips, A. J., Arauco-Brown, R., Oceguera-Figueroa, A., Gomez, G. P., Beltrán, M., Lai, Y.-T., and Siddall, M. E. 2010. Tyrannobdella rex n. gen. n. sp. and the evolutionary origins of mucosal leech infestations. *PLoS ONE* 5:e10057+.

Piel, W. H., Donoghue, M., and Sanderson, M. 2002. TreeBASE: A database of phylogenetic information. *Research Report from the National Institute for Environmental Studies* 171:41-47.

Pine, R. H., Timm, R. M., and Weksler, M. 2012. A newly recognized clade of trans-Andean oryzomyini (Rodentia: Cricetidae), with description of a new genus. *Journal of Mammalogy* 93:851-870.

Pinna, M.C.C. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7:367–394.

Pirie, M. D., Humphreys, A. M., Galley, C., Barker, N. P., Verboom, G. A., Orlovich, D., Draffin, S. J., Lloyd, K., Baeza, C. M., Negritto, M., Ruiz, E., Sanchez, J. H. C., Reimer, E., and Linder, H. P. 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Molecular Phylogenetics and Evolution* 48:1106-1119.

Pisani, D. 2002. Comparing and combining trees and data in phylogenetic analysis. PhD thesis. University of Bristol.

Pisani, D., Benton, M., and Wilkinson, M. 2007. Congruence of morphological and molecular phylogenies. *Acta Biotheoretica* 55:269-281-281.

Piwowar, H. A., Day, R. S., and Fridsma, D. B. 2007. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2:e308+.

Piwowar, H. A., Vision, T. J., and Whitlock, M. C. 2011. Data archiving is a good investment. *Nature* 473:285.

Piwowar, H. A., Vision, T.J. 2013. Data reuse and the open data citation advantage. *PeerJ* 1:e175

Piwowar, H. 2013. Altmetrics: Value all research products. *Nature* 493:159

Planet, P.J. 2006 Tree disagreement: Measuring and testing incongruence in phylogenies. *Journal of Biomedical Informatics* 39:86-102.

Planet, P. J. and Sarkar, I.N. 2005. mILD: a tool for constructing and analyzing matrices of pairwise phylogenetic character incongruence tests. *Bioinformatics* 21:4423–4424.

Poe, S. 1996. Data set incongruence and the phylogeny of crocodilians. *Systematic Biology* 45:393–414.

Poe, S., and Wiens, J. J. 2000. Character selection and the methodology of morphological phylogenetics pp 20–36. In: J. J. Wiens (ed.) Phylogenetic analysis of morphological data. Smithsonian Institution Press, Washington, DC.

Pogue, M.G. and Mickevich, M.F. 1990. Character definitions and character state delineation: The bête noire of phylogenetic inference. *Cladistics* 6:319–361.

Pol, D. and Escapa, I. H. 2009. Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics* 25:515-527.

Ponder, W.F., Clark, S.A, Eberhard, S., Studdert, J.B. A radiation of hydrobiid snails in the caves and streams at Precipitous Bluff, southwest Tasmania, Australia (Mollusca: Caenogastropoda: Rissooidea: Hydrobiidae s.l.). *Zootaxa* 1074:1-66.

Poyato-Ariza, F. J. 2003. Dental characters and phylogeny of pycnodontiform fishes. *Journal of Vertebrate Paleontology* 23:937-940.

Poyato-Ariza, F. J., Grande, T., and Diogo, R. 2010. Gonorynchiform interrelationships: Historic overview, analysis, and revised systematics of the group. In: Gonorynchiformes and Ostariophysan Relationships A Comprehensive Review pp. 227-337.

Pradel, A., Tafforeau, P., Maisey, J. G., and Janvier, P. 2011. A new Paleozoic Symmoriiformes (Chondrichthyes) from the Late Carboniferous of Kansas (USA) and cladistic analysis of early chondrichthyans. *PLoS ONE* 6:e24938+.

Prager, E.M. and Wilson, A.C. 1988. Ancient origin of lactalbumin from lysozyme: Analysis of DNA and amino acid sequences. *Journal of Molecular Evolution* 27:326–335.

Prentice, T. R. and Redak, R. A. 2012. *Esophyllas*, a new genus of erigonine spiders from southern California (Araneae: Linyphiidae: Erigoninae). *Zootaxa* 3265:1-21.

Pretti, V. Q., Calcagnotto, D., Toledo-Piza, M., and de Almeida-Toledo, L. F. 2009. Phylogeny of the neotropical genus *Acestrorhynchus* (Ostariophysi: Characiformes) based on nuclear and mitochondrial gene sequences and morphology: a total evidence approach. *Molecular Phylogenetics and Evolution* 52:312-320.

Prevosti, F. J. 2010. Phylogeny of the large extinct south american canids (Mammalia, Carnivora, Canidae) using a ” total evidence” approach. *Cladistics* 26:456-481.

Prideaux, G. J. and Warburton, N. M. 2010. An osteology-based appraisal of the

phylogeny and evolution of kangaroos and wallabies (Macropodidae: Marsupialia). *Zoo. J. Linn. Soc.* 159:954-987.

Puértolas, E., Canudo, J.I. and Cruzado-Caballero, P. 2011. A new crocodylian from the late Maastrichtian of Spain: Implications for the initial radiation of crocodyloids. *PLoS ONE* 6: e20011+.

Pujos, F., de Iuliis, G., Argot, C. and Werdelin, L. 2007. A peculiar climbing Megalonychidae from the Pleistocene of Peru and its implication for sloth history. *Zoological Journal of the Linnean Society* 149:179–235.

Quicke, D.L., Jones, O.R. and Epstein, D.R. 2007. Correcting the problem of false incongruence due to noise imbalance in the incongruence length difference (ILD) test. *Systematic Biology* 56:496–503.

Ragan, M. 1992. Matrix representation in reconstructing phylogenetic relationships among the eukaryotes. *Biosystems* 28:47–55.

Ragan, M.A., McInerney, J.O. and Lake, J.A. 2009. The network of life: Genome beginnings and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:2169–2175.

Ragsdale, E. J. and Baldwin, J. G. 2010. Resolving phylogenetic incongruence to articulate homology and phenotypic evolution: a case study from nematoda. *Proceedings of the Royal Society B: Biological Sciences* 277:1299-1307.

Ramirez, M.J. 2006. Further problems with the incongruence length difference test: “hypercongruence” effect and multiple comparisons. *Cladistics* 22:289–295.

Robinson, D.R. and Foulds, L.R. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53:131–147.

Rodrigo A. G., Kelly-Borges, M., Bergquist, P.R. and Bergquist, P.L. 1993. A randomization test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *New Zealand Journal of Botany* 31:257-268.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., and Rasnitsyn, A. P. 2012. A Total-Evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Systematic Biology* 61:973-999.

Rossello, F. and Valiente, G. 2006. The transposition distance for phylogenetic trees. arXiv:q-bio/0604024v1 [q-bio.PE]

Ruta, M. and Coates, M.I. 2007. Dates, nodes and character conflict: Addressing the lissamphibian origin problem. *Journal of Systematic Palaeontology* 5:69–122.

Sánchez-Villagra, M.R. and Williams, B.A. 1998. Levels of homoplasy in the evolution of the mammalian skeleton. *Journal of Mammalian Evolution* 5:113–126.

Sánchez-Villagra, M.R., Horovitz, I. and Motokawa, M. 2006. A comprehensive morphological analysis of talpid moles (Mammalia) phylogenetic relationships. *Cladistics* 22:59–88.

Sanderson, M. J. and Donoghue, M. J. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43:1781-1795.

Sang, T. 1995. New measurements of distribution of homoplasy and reliability of parsimonious cladograms. *Taxon* 44: 77-82.

Sansom, R.S., Gabbott, S.E. and Purnell, M.A. 2010. Non-random decay of chordate

characters causes bias in fossil interpretation. *Nature* 463:797-800.

Schaeffer, B., Hecht, B.M., and Eldredge, N. 1972. Phylogeny and paleontology, pp. 31-46. In: T. Dobzhansky, B.M. Hecht, and W. Steere (eds.), Evolutionary Biology vol 6, Springer US.

Schliep, K. P. 2011. Phangorn: phylogenetic analysis in R. *Bioinformatics* 27:592-593.

Schuh, R. T., Weirauch, C., and Wheeler, W. C. 2009. Phylogenetic relationships within the cimicomorpha (hemiptera: Heteroptera): a total-evidence analysis. *Systematic Entomology* 34:15-48.

SciVerse, 2013. SciVerse Titles List http://files.sciverse.com/documents/xlsx/title_list.xlsx

Scotland, R. W., Olmstead, R. G., and Bennett, J. R. 2003. Phylogeny reconstruction: The role of morphology. *Systematic Biology* 52:539-548.

Seiffert, E. 2007. A new estimate of afrotherian phylogeny based on simultaneous analysis of genomic, morphological, and fossil evidence. *BMC Evolutionary Biology* 7:224+.

Sereno, P.C. and Brusatte, S.L. 2008. Basal abelisaurid and carcharodontosaurid theropods from the Lower Cretaceous Elrhaz formation of Niger. *Acta Palaeontologica Polonica* 53:15–46.

Sharkey, M.J. 2005. A new species of Agathirsia Westwood (Hymenoptera: Braconidae: Agathidinae) from Mexico. *Zootaxa* 1070:43-47.

Shimada, K. 2005. Phylogeny of lamniform sharks (Chondrichthyes: Elasmobranchii) and the contribution of dental characters to lamniform systematics. *Paleontological Research* 9:55–72.

Shimodaira, H. and Hasegawa, M. 1999. Multiple comparisons of Log-Likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16:1114+.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51:492–508.

Siddall, M. 1998. The follies of ancestor worship. *Nature Debates*. http://www.nature.com/nature/debates/fossil/fossil_3.html

Siddall, M. E. 1997. Prior agreement: Arbitration or arbitrary? *Systematic Biology* 46:765+.

Siegel, S. and Castellan, N.J. 1998. Nonparametric Statistics for the Behavioural Sciences. McGraw-Hill, New York.

Sigwart, J. D. and M. D. Sutton 2007. Deep molluscan phylogeny: synthesis of palaeontological and neontological data. *Proceedings of the Royal Society B: Biological Sciences* 274:2413-2419.

Simmons, N.B. and Conway, T.M. 2001. Phylogenetic relationships of mormoopid bats (Chiroptera: Mormoopidae) based on morphological data. *Bulletin of the American Museum of Natural History* 258:1–100.

Simmons, N.B., Seymour, K.L., Habersetzer, J. and Gunnell, G.F. 2008. Primitive early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature* 451:818–821.

Simpson, G.G. (1937) Patterns of Phylectic Evolution. *Bull. Geol. Soc. Amer.* 48:308-309.

Simpson, G.G. (1944) Tempo and Mode in Evolution. New York: Columbia Univ. Press.

Skutschas, P. P., and Y. M. Gubin. 2012. A New Salamander from the Late Paleocene -

Early Eocene of Ukraine. *Acta Palaeontologica Polonica*. 57:135-148

Smith, A. B. 1994. Systematics and the fossil record: Documenting evolutionary patterns. Blackwell Scientific, Oxford, U.K.

Smith, V. 2009. Data publication: Towards a database of everything. *BMC Research Notes* 2:113+.

Smith, N. D. 2010. Phylogenetic analysis of pelecaniformes (aves) based on osteological data: Implications for waterbird phylogeny and fossil calibration studies. *PLoS ONE* 5:e13354+.

Smith, N. D. and Turner, A. H. 2005 Morphology's Role in Phylogeny Reconstruction: Perspectives from Paleontology. *Systematic Biology* 54:166-173.

Smith, N. D. and Pol, D. 2007. Anatomy of a basal sauropodomorph dinosaur from the early Jurassic Hanson formation of Antarctica. *Acta Palaeontologica Polonica* 52:657–674.

Sokal, R.R. and Sneath, P.H.A. 1963. Principles of Numerical Taxonomy. W.H. Freeman & Co., San Francisco.

Song, H. and Bucheli, S.R. 2010. Comparison of phylogenetic signal between male genitalia and non-genital characters in insect systematics. *Cladistics* 26:23–35.

Spangler, R. E. and Olmstead, R. G. 1999. Phylogenetic analysis of Bignoniaceae based on the cpDNA gene sequences rbcL and ndhF. *Annals of the Missouri Botanical Garden* 86: 33-46.

Sparks, J.S. 2008. Phylogeny of the cichlid subfamily Etroplinae and taxonomic revision of the Malagasy cichlid genus Paretroplus (Teleostei: Cichlidae). *Bulletin of the American Museum of Natural History* 314:1+.

Spaulding, M., O'Leary, M.A. and Gatesy, J. 2009. Relationships of Cetacea (Artiodactyla) among mammals: Increased taxon sampling alters interpretations of key fossils and character evolution. *PLoS ONE* 4:e7062+.

Spaulding, M and Flynn, J. 2012. Phylogeny of the Carnivoramorpha: The impact of postcranial characters. *Journal of Systematic Palaeontology* 10:653-677.

Springer, M.S., Burk-Herrick, A., Meredith, R., Eizirik, E., Teeling, E., O'Brien, S.J. and Murphy, W.J. 2007. The adequacy of morphology for reconstructing the early history of placental mammals. *Systematic Biology* 56:673–684.

Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., Stadler, T., Steiner, C., Ryder, O. A., Janečka, J. E., Fisher, C. A., and Murphy, W. J. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS ONE* 7:e49521+.

Stackoverflow, 2012 How do I grep in parallel? http://stackoverflow.com/questions/11898949/how-do-i-grep-in-parallel?answertab=active#tab-top

Steel, M.A. and Penny, D. 1993. Distributions of tree comparison metrics – some new results. *Systematic Biology* 42:126–141.

Stoltzfus, A., O'Meara, B., Whitacre, J., Mounce, R., Gillespie, E., Kumar, S., Rosauer, D., and Vos, R. 2012. Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes* 5:574+.

Sues, H.-D. and Averianov, A. 2009. A new basal hadrosauroid dinosaur from the Late Cretaceous of Uzbekistan and the early radiation of duck-billed dinosaurs. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 276:2549–2555.

Sues, H.-D. and Reisz, R.R. 2008. Anatomy and phylogenetic relationships of Sclerosaurus armatus (Amniota: Parareptilia) from the Buntsandstein (Triassic) of Europe. *Journal of Vertebrate Paleontology* 28:1031–1042.

Sullivan J., 1996 Combining data with different distributions of among-site rate variation *Systematic Biology* 45:375-380

Swofford, D.L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D.L., Olsen, G. J., Waddell, P. G. and Hillis, D. M. 1996. Phylogenetic inference. In: D.M. Hillis, C. Moritz and B.K. Mable, editors. Molecular Systematics. Sinauer Associates. Sunderland, Massachusetts: p. 407–514.

Tange, O. 2011. GNU Parallel - The Command-Line Power Tool. *Login: The USENIX Magazine* Feb:42-47.

Templeton, A.R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37:221-244.

Thomson, R. C. and Shaffer, H. B. 2010. Sparse supermatrices for phylogenetic inference: Taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Systematic Biology* 59:42-58.

Thomson Reuters, 2013. Web of Knowledge: Who We Are http://wokinfo.com/about/whoweare/

Thorley, J.L. and Page, R.D.M. 2000. RadCon: Phylogenetic tree comparison and consensus. *Bioinformatics* 16:486-487.

Triapitsyn, S. V., Vickerman, D.B., Heraty, J. M., Logarzo, G.A. 2006. A new species of Gonatocerus (Hymenoptera: Mymaridae) parasitic on proconiine sharpshooters (Hemiptera: Cicadellidae) in the New World. *Zootaxa* 1158:55-67.

Trueb, L. and A. M. Baez 2006. Revision of the Early Cretaceous *Cordicephalus* from Israel and an assessment of its relationships among pipoid frogs. *Journal of Vertebrate Paleontology* 26:44-59.

Turner, A.H. and Sertich, J.J.W. 2010. Phylogenetic history of *Simosuchus clarki* (Crocodyliformes: Notosuchia) from the late cretaceous of madagascar. *Journal of Vertebrate Paleontology* 30: 177-236.

Vallin, G. and Laurin, M. 2004. Cranial morphology and affinities of *Microbrachis*, and a reappraisal of the phylogeny and lifestyle of the first amphibians. *Journal of Vertebrate Paleontology* 24:56–72.

van der Linde, K., Houle, D., Spicer, G. S., and Steppan, S. J. 2010. A supermatrix-based molecular phylogeny of the family drosophilidae. *Genetics Research* 92:25-38.

Vea, I. M. and D. A. Grimaldi 2012. Phylogeny of ensign scale insects (Hemiptera: Coccoidea: Ortheziidae) based on the morphology of recent and fossil females. *Systematic Entomology* 37:758-783.

Venczel, M. 2008. A new salamandrid amphibian from the Middle Miocene of Hungary and

its phylogenetic relationships. *Journal of Systematic Palaeontology* 6:41–59.

Vermeij, G.J. and Carlson, S.J. 2000. The muricid gastropod subfamily Rapaninae: Phylogeny and ecological history. *Paleobiology* 26:19–46.

Vidal, N., and Lecointre, G. 1998. Weighting and congruence: A case study based on three mitochondrial genes in pitvipers. *Molecular Phylogenetics and Evolution* 9: 366 –374.

Vieira, G.H.C., Colli, G.R. and Bao, S.N. 2005. Phylogenetic relationships of corytophanid lizards (Iguania, Squamata, Reptilia) based on partitioned and total evidence analyses of sperm morphology, gross morphology, and DNA data. *Zoologica Scripta* 34:605–625.

Vilhelmsen, L., Perrichot, V., and Shaw, S. R. 2010. Past and present diversity and distribution in the parasitic wasp family megalyridae (Hymenoptera). *Systematic Entomology* 35:658-677.

von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell'Ampio, E., Pass, G., Ebersberger, I., Meyer, B., Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K. and Misoff, B. 2012. Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Molecular Biology and Evolution* 29: 1031-1045.

W3C, 1991. http://www.w3.org/Protocols/HTTP/AsImplemented.html

Walters, W. H. 2011. Comparative Recall and Precision of Simple and Expert Searches in Google Scholar and Eight Other Databases. *Portal: libraries and the academy* 11:971-1006.

Ward, C.V. 1997. Functional anatomy and phyletic implications of the hominoid trunk and hindlimb pp. 101–130. In: B. Begun, C. Ward, and M. Rose (eds.) Function, Phylogeny, and Fossils: Miocene Hominoid Evolution and Adaptations. Plenum Press, New York.

Waterman, M.S. and Smith, T.F. 1978. On the similarity of dendrograms. *Journal of Theoretical Biology* 73:789–800.

Webb, D.W. 2005. New Genera of Neotropical Therevidae (Insecta: Diptera). *Zootaxa* 1091:1-26.

Weksler, M. 2006. Phylogenetic relationships of oryzomine rodents (Muroidea: Sigmodontinae): Separate and combined analyses of morphological and molecular data. *Bulletin of the American Museum of Natural History* 296:1+.

Wheeler, Q.D. 1981. The ins and outs of character analysis: A response to Crisci and Stuessy. *Systematic Botany* 6: 297-306.

Wheeler, W C 1991. Congruence among data sets: A Bayesian Approach pp. 334 – 346. In: M M Miyamoto and J Cracraft (eds.) Phylogenetic analysis of DNA sequences. Oxford University Press

Wheeler, W. C. 1999. Measuring topological congruence by extending character techniques. *Cladistics* 15:131–135.

White, E.P., Baldridge, E., Brym, Z.T. Locey, K.J., McGlinn, D.J. and Supp, S.R. 2013. Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*. 6: 1-10.

Whitlock, J. A. 2010. Phylogenetic relationships of the Eocene percomorph fishes †*Priscacara* and †*Mioplosus*. *Journal of Vertebrate Paleontology* 30:1037-1048.

Wible, J. R., Rougier, G.W., Novacek, M.J. and Asher, R.J. 2009. The eutherian mammal

Maelestes gobiensis from the Late Cretaceous of Mongolia and the phylogeny of Cretaceous Eutheria. *Bulletin of the American Museum of Natural History* 327:1–123.

Wickham, H. 2011. The Split-Apply-combine strategy for data analysis. *Journal of Statistical Software* 40:1-29.

Wiens, J. J. 2003a. Incomplete taxa, incomplete characters, and phylogenetic accuracy: Is there a missing data problem? *Journal of Vertebrate Paleontology* 23:297-310.

Wiens, J. J. 2003b. Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology* 52:528-538.

Wiens, J. J. 2004. The role of morphological data in phylogeny reconstruction. *Systematic Biology* 53:653-661.

Wiens, J. J. & Hillis, D.M. 1996. Accuracy of parsimony analysis using morphological data: a reappraisal. 21:237-253.

Wiens, J. J. and Hollingsworth, B. D. 2000. War of the iguanas: Conflicting molecular and morphological phylogenies and Long-Branch attraction in iguanid lizards. Systematic biology 49:143-159.

Wiens, J., Fetzner, J., Parkinson, C. and Reeder, T. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. *Systematic Biology* 54:719–748.

Wilkinson, M. 1995a. Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology* 44:501-514.

Wilkinson, M. 1995b. TAXEQ3: safe taxonomic reduction using taxonomic equivalence. Computer program published by the author. http://www.nhm.ac.uk/research-curation/research/projects/software/

Wilkinson, M. 1997. Characters, congruence and quality: A study of Neuroanatomical and Traditional Data in Caecilian Phylogeny. *Biological Reviews* 72: 423-470.

Wilkinson, M. 2001 PICA 4.0: software and documentation. Department of Zoology, The Natural History Museum, London. http://www.nhm.ac.uk/research-curation/research/projects/software/

Williams, W.T. and Clifford, H.T. 1971. On the comparison of two classifications of the same set of elements. *Taxon* 20: 519-522.

Wills, M.A. 1999. Congruence between phylogeny and stratigraphy: Randomization tests and the gap excess ratio. *Systematic Biology* 48:559–580.

Wilson, E.O. 1998. *Consilience: The Unity of Knowledge*. Alfred A. Knopf, New York.

Wilson, J. J. 2010. Assessing the value of DNA barcodes and other priority gene regions for molecular phylogenetics of lepidoptera. *PLoS ONE* 5:e10525+.

Winther, R. 2009. Character analysis in cladistics: Abstraction, reification, and the search for objectivity. *Acta Biotheoretica* 57:129–162–162.

Wolsan, M. and Sato, J. J. 2010. Effects of data incompleteness on the relative performance of parsimony and bayesian approaches in a supermatrix phylogenetic reconstruction of Mustelidae and Procyonidae (Carnivora). *Cladistics* 26:168-194.

Wood, H. M., Griswold, C. E., and Gillespie, R. G. 2012. Phylogenetic placement of pelican spiders (Archaeidae, Araneae), with insight into evolution of the ” neck” and predatory behaviours of the superfamily Palpimanoidea. *Cladistics* 28:598-626.

Worthy, T.H. 2009. Descriptions and phylogenetic relationships of two new genera and four new species of Oligo-Miocene waterfowl (Aves: Anatidae) from Australia. *Zoological Journal of the Linnean Society* 156:411–454.

Worthy, T and Scofield 2012. Twenty-first century advances in knowledge of the biology of moa (Aves:Dinornithiformes): a new morphological analysis and moa diagnoses revised. *New Zealand Journal of Zoology*, 39:87-153.

Yoder, A. D., Irwin, J. A., and Payseur, B. A. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. Systematic biology 50:408-424.

Young, M.T. and de Andrade. M.B. 2009. What is *Geosaurus*?: redescription of *Geosaurus giganteus* (Thalattosuchia: Metriorhynchidae) from the Upper Jurassic of Bayern, Germany. *Zoological Journal of the Linnean Society* 157:551–585.

Young, N.M. 2005. Estimating hominoid phylogeny from morphological data: character choice, phylogenetic signal and postcranial data pp 19-31. In: D.E. Lieberman, R.J. Smith, J. Kelley (eds.) Interpreting the Past: Essays on Human, Primate and Mammal Evolution in Honor of David Pilbeam. Brill Academic Publishers. Boston.

Zanno, L.E., Gillette, D.D., Albright, L.B. and Titus, A.L. 2009. A new North American therizinosaurid and the role of herbivory in 'predatory' dinosaur evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 276:3505-3511.

Zaretskii, K.A. 1965. Constructing a tree on the basis of a set of distances between the hanging vertices. *Uspekhi Mat. Nauk* 20:90–92.

Zhang, J. Y. 2006. Phylogeny of Osteoglossomorpha. *Vertebrata PalAsiatica* 44:43-59.

Zrzavy, J., Riha, P., Pialek, L., and Janouskovec, J. 2009. Phylogeny of Annelida (Lophotrochozoa): total-evidence analysis of morphology and six genes. *BMC Evolutionary Biology* 9:189+.

Appendices

Appendix 2.1

(on next pages) Summary statistics for the 62 vertebrate morphological data sets analysed herein. Data set dimensions (numbers of taxa and informative characters) refer to the pre-processed matrices after applying safe taxonomic deletion rules (see text for details). ILD column reports the p-value resulting from an incongruence length difference test with 999 random partitions. TILD column reports p-value resulting from a topological incongruence length difference test (999 random partitions). IRD columns report the p-values resulting from incongruence relationship difference tests with 99 random partitions. IRD_{NND+RF} denotes the IRD test using the Robinson Foulds tree-to-tree distance for nearest neighbouring trees. IRD_{MR+RF} denotes the IRD test using the Robinson Foulds tree-to-tree distance for majority rule trees. $IRD_{MR+AgD1}$ denotes the IRD test using the maximum agreement subtree tree-to-tree distance for majority rule trees. CI and RI columns give ensemble consistency and retention indices respectively for entire data sets. pCI and pRI columns give ensemble consistency and retention indices (respectively) for partitions of the data set (cranial or postcranial). Mean ci and Mean ri give the mean per character consistency and retention indices (respectively) for partitions of the data set (cranial or postcranial). MWU p-value gives the p-value from a Mann-Whitney U test of cranial and postcranial per character ci values *within* each data set. pHER gives the homoplasy excess ratio for partitions of the data set (cranial or postcranial) derived from 999 randomized matrices.

First author	Year	Group	Broad Group	#Taxa	# Chars.	# Cranial Chars.	# Postcranial Chars.	Cranial MPTs	Postcranial MPTs	ILD	TILD	IRD_{NND+RF}	$IRD_{MRs\ RF}$	$IRD_{MRs\ AgD1}$	CI (whole)	RI (whole)	CI (cranial)	CI (postcranial)	mean ci (cranial)	mean ci (postcranial)	MWU p-value	RI (cranial)	RI (postcranial)	mean ri (cranial)	mean ri (postcranial)	HER (cranial)	HER (postcranial)
Allain	2008	Sauropods	Ornithodira	18	212	62	150	6	3	0.006	0.001	0.040	0.070	0.350	0.670	0.764	0.683	0.686	0.691	0.767	0.731	0.769	0.137	0.794	0.774	0.662	0.682
Anderson	2008	Batrachia	Other tetrapods	52	219	143	76	1052	54	0.001	0.001	0.100	0.240	0.110	0.254	0.581	0.254	0.285	0.351	0.372	0.525	0.556	0.195	0.595	0.622	0.407	0.379
Asher	2005	Lagomorpha	Mammals	29	228	170	58	2	1	0.001	0.001	0.020	0.010	0.210	0.349	0.610	0.373	0.331	0.449	0.401	0.588	0.572	0.232	0.640	0.652	0.501	0.429
Asher	2006	Afrotheria	Mammals	23	126	96	30	1	>10000	0.781	0.001	0.190	0.210	0.160	0.402	0.611	0.406	0.441	0.479	0.439	0.573	0.572	0.888	0.614	0.600	0.404	0.325
Asher	2007	Eutheria	Mammals	46	196	121	75	191	1	0.001	0.001	0.020	0.070	0.470	0.242	0.527	0.223	0.275	0.309	0.413	0.483	0.621	0.017	0.522	0.614	0.286	0.384
Beard	2009	Amphipithecidae	Mammals	39	326	252	64	7	>10000	0.845	0.001	0.160	0.010	0.010	0.306	0.547	0.305	0.569	0.400	0.639	0.491	0.636	0.000	0.545	0.564	0.426	0.427
Beck	2008	Marsupiala	Mammals	27	245	103	139	4	8	0.004	0.001	0.220	0.110	0.290	0.378	0.651	0.432	0.400	0.543	0.508	0.650	0.656	0.349	0.696	0.633	0.543	0.477
Bloch	2007	Plesiadapiforms	Mammals	21	173	107	65	3	>10000	0.103	0.001	0.310	0.170	0.130	0.443	0.549	0.401	0.627	0.444	0.699	0.480	0.642	0.000	0.556	0.521	0.337	0.295
Bourdon	2009	Palaeognathae	Ornithodira	17	129	35	94	3	2	1.000	1.000	0.660	0.880	0.570	0.882	0.951	0.956	0.855	0.743	0.821	0.753	0.855	0.593	0.983	0.938	0.975	0.921
Butler	2008	Ornithischia	Ornithodira	46	218	130	88	>10000	>10000	0.980	0.001	0.020	0.020	0.130	0.503	0.732	0.502	0.521	0.603	0.614	0.691	0.714	0.774	0.721	0.743	0.574	0.604
Carrano	2008	Ceratosauria	Ornithodira	21	151	71	80	>10000	7211	1.000	1.000	0.010	0.020	0.140	0.741	0.816	0.796	0.685	0.779	0.685	0.772	0.629	0.051	0.881	0.752	0.778	0.586
Ezcurra	2007	Coelophysoidea	Ornithodira	11	136	62	74	1	6	0.011	0.080	0.010	0.010	0.140	0.663	0.720	0.610	0.731	0.678	0.728	0.684	0.735	0.216	0.698	0.785	0.617	0.668
Friedman	2007	Actinistia	Fishes	39	195	146	49	1320	>10000	0.005	0.001	0.650	0.140	0.290	0.453	0.709	0.464	0.523	0.569	0.583	0.706	0.715	0.422	0.716	0.689	0.580	0.511
Friedman	2008	Pleuronectiformes	Fishes	19	58	16	42	43	12	0.215	0.001	0.430	0.460	0.270	0.507	0.732	0.548	0.526	0.380	0.526	0.487	0.676	0.121	0.655	0.754	0.423	0.608
Frobisch	2007	Dicynodontia	Other tetrapods	42	99	86	13	30	>10000	0.869	0.001	0.180	0.040	0.040	0.492	0.764	0.468	0.567	0.572	0.462	0.748	0.631	0.567	0.769	0.712	0.659	0.078
Gaffney	2009	Bothremydidae	Other tetrapods	47	174	123	51	>10000	>10000	0.482	0.001	0.010	0.010	0.010	0.686	0.805	0.558	0.686	0.678	0.664	0.802	0.721	0.936	0.818	0.762	0.726	0.614
Gates	2007	Hadrosaurinae	Ornithodira	15	120	61	34	4	8354	0.171	0.007	0.090	0.130	0.030	0.579	0.660	0.585	0.750	0.641	0.750	0.641	0.736	0.203	0.682	0.444	0.524	0.354
Gaudin	2009	Pholidota	Mammals	17	388	298	100	2	1922	0.540	0.014	0.010	0.100	0.200	0.515	0.646	0.339	0.325	0.570	0.678	0.589	0.655	0.001	0.637	0.685	0.522	0.532
Gaulbert	2005	Feliformia	Mammals	39	329	229	100	28	27	0.001	0.001	0.010	0.010	0.460	0.343	0.619	0.356	0.335	0.454	0.381	0.561	0.626	0.059	0.620	0.666	0.497	0.519
Godefroit	2006	Titanosauria	Ornithodira	21	56	43	13	5	661	1.000	1.000	0.380	0.360	0.170	0.906	0.970	0.893	1.000	0.723	0.723	0.741	0.741	0.000	0.965	1.000	0.935	1.000
Gonzalez-Riga	2009	Hadrosauridae	Ornithodira	23	84	15	69	>10000	20	0.998	0.001	0.370	0.560	0.030	0.575	0.678	0.850	0.549	0.378	0.607	0.333	0.632	0.046	0.762	0.670	0.105	-0.207
Hill	2005	Amniota	Other tetrapods	80	345	162	183	>10000	>10000	0.007	0.001	0.030	0.110	0.060	0.245	0.732	0.215	0.298	0.295	0.436	0.658	0.744	0.000	0.702	0.768	0.568	0.654
Hilton	2009	Acipenseriformes	Fishes	16	48	31	17	2	46	0.664	0.001	0.200	0.050	0.430	0.634	0.790	0.634	0.724	0.661	0.529	0.644	0.413	0.257	0.780	0.839	0.670	0.739
Holland	2009	Tetrapodomorpha	Other tetrapods	12	103	54	45	1	1	0.008	0.009	0.110	0.490	0.390	0.722	0.785	0.800	0.710	0.722	0.667	0.702	0.615	0.338	0.826	0.745	0.777	0.655
Hospitaleche	2007	Sphenisciformes	Ornithodira	22	44	20	23	25	>10000	0.935	0.035	0.820	0.300	0.280	0.436	0.670	0.442	0.467	0.439	0.525	0.607	0.605	0.295	0.706	0.626	0.447	0.270
Hurley	2007	Actinopterygii	Fishes	29	70	54	16	42	1136	0.586	0.001	0.910	0.870	0.630	0.470	0.702	0.483	0.509	0.603	0.579	0.704	0.748	0.977	0.725	0.640	0.531	0.359
Inamura	2005	Cottoidei	Fishes	8	44	23	21	2	3	0.405	0.001	0.250	0.980	0.580	0.660	0.667	0.640	0.700	0.469	0.738	0.302	0.643	0.052	0.542	0.778	0.390	0.601
Jouve	2006	Crocodylomorpha	Other tetrapods	45	217	190	27	120	>10000	0.398	0.177	0.430	0.490	0.010	0.344	0.644	0.340	0.526	0.486	0.504	0.642	0.616	0.429	0.646	0.626	0.509	0.324
Ksepka	2009	Galliformes	Ornithodira	64	120	28	89	>10000	8472	0.079	0.177	0.040	0.030	0.610	0.330	0.781	0.433	0.342	0.565	0.523	0.772	0.759	0.527	0.780	0.785	0.645	0.688
Lee	2002	Serpentes	Other tetrapods	23	263	204	57	3	32	0.629	0.001	0.660	0.040	0.130	0.471	0.670	0.485	0.529	0.600	0.545	0.655	0.632	0.700	0.669	0.669	0.538	0.490
Li	2007	Squamata	Other tetrapods	23	399	263	136	1	12	0.099	0.005	0.450	0.530	0.410	0.555	0.672	0.523	0.538	0.627	0.638	0.638	0.650	0.650	0.672	0.669	0.567	0.548
Lister	2005	Cervidae	Mammals	10	62	21	41	6	4	0.206	0.001	0.350	0.370	0.250	0.647	0.544	0.750	0.623	0.504	0.580	0.353	0.417	0.914	0.605	0.519	0.418	0.261
Lu	2009	Pterosauria	Ornithodira	57	117	62	55	>10000	>10000	0.096	0.041	0.010	0.280	0.010	0.436	0.798	0.510	0.423	0.666	0.547	0.829	0.780	0.057	0.827	0.781	0.736	0.648
Lyson	2009	Baenidae	Other tetrapods	16	54	34	19	43	>10000	0.990	0.001	0.320	0.010	0.520	0.558	0.688	0.550	0.655	0.543	0.500	0.589	0.500	1.000	0.686	0.744	0.504	0.354
Martinelli	2007	Ictidosauria	Other tetrapods	22	85	68	17	2	>10000	0.161	0.004	0.060	0.330	0.190	0.643	0.805	0.629	0.857	0.701	0.500	0.753	0.480	0.109	0.800	0.833	0.686	0.765
Martinez	2009	Dinosauria	Ornithodira	12	98	28	70	45	3	0.071	0.001	0.640	0.050	0.380	0.561	0.568	0.732	0.560	0.577	0.597	0.554	0.492	0.944	0.741	0.500	0.544	0.251
Matsumoto	2009	Choristodera	Other tetrapods	14	81	49	29	2	10	0.002	0.109	0.430	0.660	0.210	0.582	0.690	0.647	0.586	0.658	0.509	0.681	0.519	0.045	0.760	0.688	0.611	0.422
Muller	2006	Eureptiles	Other tetrapods	25	90	71	19	3	>10000	0.189	0.146	0.050	0.670	0.080	0.415	0.649	0.429	0.468	0.555	0.474	0.661	0.600	0.808	0.657	0.624	0.467	0.345
Osi	2009	Ankylosauria	Ornithodira	18	58	44	14	4	>10000	0.662	0.001	0.010	0.120	0.210	0.483	0.672	0.464	0.643	0.537	0.471	0.667	0.430	0.774	0.691	0.571	0.465	0.147
Parenti	2008	Adrianichthyidae	Fishes	31	80	37	43	3818	8	0.588	0.001	0.010	0.090	0.180	0.595	0.814	0.677	0.584	0.576	0.549	0.666	0.590	0.781	0.867	0.789	0.778	0.688
Phillips	2009	Monotremes	Mammals	96	439	328	111	270	>10000	0.830	0.001	0.050	0.070	0.020	0.381	0.784	0.337	0.503	0.464	0.594	0.747	0.841	0.000	0.760	0.875	0.682	0.811
Pujos	2007	Pholidota	Mammals	18	42	25	17	4	21	0.072	0.001	0.630	0.010	0.730	0.472	0.639	0.455	0.628	0.417	0.525	0.563	0.589	0.096	0.646	0.625	0.382	0.457
Ruta	2007	Lissamphibia	Other tetrapods	46	333	233	100	12	240	0.001	0.001	0.090	0.130	0.010	0.343	0.624	0.320	0.367	0.450	0.481	0.550	0.650	0.918	0.630	0.672	0.460	0.463
Sanchez-Villagra	2006	Talpidae	Mammals	17	157	74	83	4	1	0.003	0.008	0.070	0.120	0.180	0.515	0.674	0.495	0.551	0.587	0.598	0.596	0.668	0.690	0.629	0.730	0.445	0.609
Sereno	2008	Carcharodontosaurids	Ornithodira	9	60	37	23	3	18	0.498	0.036	0.100	0.420	0.680	0.738	0.738	0.848	0.639	0.743	0.558	0.689	0.413	0.033	0.860	0.588	0.950	0.201
Shimada	2005	Lamniformes	Fishes	17	61	43	18	15	3	0.382	0.021	0.690	0.940	0.120	0.525	0.717	0.525	0.810	0.568	0.489	0.574	0.542	0.472	0.659	0.917	0.551	0.860
Simmons	2001	Mormoopidae	Mammals	17	209	103	106	2	419	1.000	0.262	0.270	0.820	0.990	0.622	0.751	0.650	0.659	0.656	0.704	0.673	0.753	0.357	0.758	0.746	0.679	0.685
Simmons	2008	Chiroptera	Mammals	29	207	48	159	24	6	0.102	0.001	0.150	0.080	0.130	0.398	0.620	0.445	0.419	0.495	0.521	0.647	0.586	0.695	0.645	0.605	0.525	0.459
Smith	2007	Plateosauria	Ornithodira	47	361	117	244	>10000	>10000	0.810	0.001	0.020	0.030	0.200	0.376	0.695	0.461	0.362	0.536	0.468	0.627	0.640	0.859	0.681	0.690	0.563	0.581
Sparks	2008	Etroplinae	Fishes	25	82	25	57	1	1	0.593	0.133	0.670	0.820	0.450	0.737	0.942	0.867	0.707	0.600	0.815	0.652	0.883	0.046	0.970	0.926	0.962	0.925
Spaulding	2009	Artiodactyla	Mammals	37	661	435	177	10	5	0.018	0.001	0.030	0.020	0.110	0.269	0.510	0.235	0.322	0.320	0.441	0.431	0.584	0.000	0.478	0.618	0.288	0.443
Sues	2008	Parareptilia	Other tetrapods	12	51	34	15	1	117	0.030	0.007	0.010	0.010	0.280	0.691	0.792	0.731	0.750	0.716	0.500	0.703	0.333	0.161	0.843	0.500	0.738	0.631
Sues	2009	Iguanodontidae	Ornithodira	34	138	100	38	72	>10000	0.840	0.001	0.090	0.060	0.080	0.544	0.861	0.528	0.651	0.630	0.689	0.801	0.773	0.307	0.859	0.870	0.784	0.798
Vallin	2004	Stegocephalia	Other tetrapods	49	159	99	60	52	448	0.001	0.001	0.030	0.030	0.010	0.359	0.710	0.355	0.417	0.447	0.527	0.706	0.716	0.094	0.718	0.701	0.584	0.544
Venczel	2008	Caudata	Other tetrapods	15	35	19	16	1	54	1.000	0.252	0.060	0.480	0.220	0.629	0.805	0.538	0.783	0.491	0.604	0.589	0.557	0.239	0.757	0.886	0.516	0.876
Vieira	2005	Corytophanidae	Other tetrapods	13	89	41	43	5	10	0.432	0.155	0.640	0.250	0.490	0.709	0.833	0.788	0.673	0.796	0.640	0.812	0.585	0.040	0.881	0.747	0.837	0.674
Weksler	2006	Oryzomyini	Mammals	54	99	57	38	6	>10000	0.562	0.001	0.200	0.050	0.070	0.230	0.620	0.275	0.388	0.386	0.459	0.614	0.580	0.086	0.625	0.604	0.567	0.434
Wible	2009	Eutheria	Mammals	69	387	321	66	40	>10000	0.430	0.001	0.010	0.170	0.010	0.254	0.551	0.254	0.293	0.327	0.336	0.510	0.492	0.235	0.562	0.464	0.397	0.105
Wiens	2005	Hylidae	Other tetrapods	63	140	64	76	>10000	1372	0.001	0.001	0.590	0.370	0.600	0.224	0.551	0.239	0.251	0.272	0.345	0.451	0.535	0.312	0.588	0.623	0.329	0.399
Worthy	2009	Anseriformes	Ornithodira	54	150	28	122	102	388	0.001	0.001	0.050	0.090	0.100	0.225	0.598	0.250	0.234	0.377	0.304	0.630	0.575	0.457	0.629	0.624	0.314	0.453
Young	2009	Crocodylomorpha	Other tetrapods	86	166	128	38	12	>10000	1.000	0.001	0.080	0.020	0.010	0.447	0.865	0.405	0.875	0.543	0.741	0.814	0.787	0.002	0.856	0.948	0.775	0.913
Zanno	2009	Therizinosauria	Ornithodira	76	348	90	258	>10000	>10000	0.215	0.001	0.030	0.010	0.010	0.354	0.701	0.380	0.362	0.463	0.475	0.684	0.682	0.714	0.708	0.699	0.528	0.570

Appendix 3.1 Data to reproduce PAUP bug

Comparing the two trees below in PAUP* version 4.0b10 for Unix (and Windows) will reliably crash the program every single time, if they are compared using the AgD1 or AgD metrics. The message given is “Segmentation fault (core dumped)”. Sometimes an 'apparent' AgD1 tree distance is printed in output log files as an impossibly large number e.g. 4294967293

Further instruction, including the exact commands to input into PAUP* to recreate this bug are given along with the data and tree files needed to reproduce it online at: https://github.com/rossmounce/extinct_extant_chapter/tree/master/PAUP_bug

tree Strict = [&U] (Acanthostega,(Proterogyrinus,Seymouria_baylorensis),
((Balanerpeton,Dendrerpeton,Tuditanus,(Asaphestera,(Microbrachis,Adelogyrinus)),
(Hapsidoparion,Saxonerpeton),(Pantylus,Stegotretus),
((Cardiocephalus_peabodyi,Euryodus_primus),Euryodus_dalyae),
(Pelodosotis,Micraroter),Rhynchonkos,Eocaecilia,Batropetes,Utaherpeton,
((Sauropleura_scalaris,Urocordylus),Ptyonius),
(((Keraterpeton_galvani,Batrachiderpeton),Diceratosaurus,
(Diplocaulus_magnicornis,Diploceraspis)),Scincosaurus),Brachydectes,Oestocephalus,Phlegethontia,Limnoscelis,Branchiosauridae,Micromelerpetontidae,
(Ecolsonia,Acheloma,Tambachia),Eryops,Doleserpeton,Salamanders,
(Frogs,Triadobatrachus),Albanerpetontidae,Micropholis,Eoscopus,Gerobatrachus,Platyrhinops,Amphibamus),Greererpeton));

tree Strict = [&U] (Acanthostega,(Proterogyrinus,
((((((((((Tuditanus,Stegotretus),Pelodosotis),Saxonerpeton),Pantylus),Asaphestera),Batropetes),((Cardiocephalus_peabodyi,Euryodus_primus),
(Rhynchonkos,Eocaecilia))),Micraroter),Euryodus_dalyae),Utaherpeton),
(Seymouria_baylorensis,Limnoscelis))),((Balanerpeton,Dendrerpeton),
((((((Hapsidoparion,Microbrachis),Brachydectes),Adelogyrinus),((Sauropleura_scalaris,
(Ptyonius,Urocordylus,Scincosaurus),((Keraterpeton_galvani,Diceratosaurus),
(Batrachiderpeton,(Diplocaulus_magnicornis,Diploceraspis)))),

(Oestocephalus,Phlegethontia))),(Salamanders,Albanerpetontidae),
(Frogs,Triadobatrachus)),Doleserpeton,Gerobatrachus,Platyrhinops),Greererpeton,
(Branchiosauridae,Micromelerpetontidae),Ecolsonia,Acheloma,Eryops,Micropholis,Eoscop
us,Tambachia,Amphibamus));

Appendix 3.2 Extra Summary Data

Extra Summary Table for Fossil Taxa data

Identifier	Tax	Mean RF	SD RF	SE RF	Max RF	Min RF	Mean PD	SD PD	SE PD	Max PD	Min PD	Mean diff MPTs	SD diff MPTs	SE diff MPTs	Max diff MPTs	Min diff MPTs
adephaga	17	5.298	5.479	1.329	23.808	2.417	21.413	18.668	4.528	89.471	10.994	49.412	180.000	43.656	736	-16
adnet	7	6.896	6.562	2.480	15.856	0.000	18.701	16.081	6.078	40.965	0.000	36.429	58.731	22.198	138	0
ahyong06	12	4.275	3.064	0.884	9.571	1.000	12.012	7.053	2.036	23.930	3.536	11.250	31.157	8.994	109	-2
apesteg	6	1.667	4.082	1.667	10.000	0.000	4.387	10.746	4.387	26.321	0.000	-0.500	0.548	0.224	0	-1
arango7	4	38.272	18.624	9.312	52.551	10.875	173.951	97.906	48.953	259.584	33.046	1697.000	1511.345	755.673	2968	-16
archostemata	8	6.333	4.113	1.454	11.667	2.000	13.948	7.137	2.523	23.373	6.633	1.250	1.753	0.620	5	0
asher03	39	9.312	15.193	2.433	57.387	0.000	36.614	50.947	8.158	176.385	0.000	44.923	123.377	19.756	660	-4
babot	13	10.183	5.342	1.482	17.143	2.000	23.082	9.838	2.729	37.740	6.325	4.846	6.135	1.702	20	0
beard	34	3.623	3.365	0.577	13.000	0.000	12.608	10.657	1.828	38.544	0.000	1.118	11.594	1.988	57	-7
bisconti	22	3.018	1.528	0.326	6.333	1.333	11.454	4.999	1.066	23.508	5.333	1.273	4.723	1.007	18	-4
blag9	7	0.714	1.496	0.565	4.000	0.000	2.176	4.661	1.762	12.490	0.000	0.286	0.756	0.286	2	0
blago4	22	13.638	14.382	3.066	44.421	0.000	45.494	46.696	9.956	155.664	0.000	27.455	55.340	11.799	195	-4
bloch	16	3.971	3.742	0.936	12.000	0.000	8.918	7.155	1.789	23.130	0.000	3.125	3.344	0.836	10	0
boess	17	1.739	1.197	0.290	5.250	0.000	6.595	3.486	0.846	13.477	0.000	-7.824	8.017	1.945	7	-18
bouetel	15	1.969	2.499	0.645	10.000	0.143	5.214	5.083	1.313	19.780	0.527	-6.467	13.384	3.456	24	-24
bourdon	9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0
brochu	21	7.427	0.721	0.157	8.907	6.125	18.555	1.262	0.275	20.927	16.199	-61.762	191.148	41.712	432	-384
claeson	12	15.089	11.019	3.181	28.121	2.667	53.429	35.243	10.174	95.104	12.894	1114.500	1417.954	409.328	4176	0
clarke9	5	5.510	7.173	3.208	18.216	1.333	22.510	22.580	10.098	62.308	8.743	20.800	48.757	21.805	108	-2
deng	26	2.123	5.808	1.139	30.000	0.000	7.903	18.704	3.668	94.446	0.000	-22.077	11.913	2.336	0	-36
dietz	6	4.768	4.267	1.742	10.857	0.000	10.456	8.185	3.341	20.672	0.000	3.500	5.394	2.202	13	0
diogo7	7	1.683	1.440	0.544	4.625	0.313	12.416	13.793	5.213	42.500	2.050	18.286	48.379	18.286	128	0
diogo8	5	2.333	0.816	0.365	3.333	1.500	11.623	2.530	1.132	14.566	8.890	0.000	0.000	0.000	0	0
evanidae	16	4.841	2.835	0.709	10.667	1.000	11.329	5.465	1.366	20.740	3.082	7.438	16.621	4.155	54	-4
fika	10	10.390	6.595	2.086	22.314	3.224	27.863	12.642	3.998	47.304	13.210	549.000	923.955	292.180	1932	-282
fordyce	15	2.267	1.580	0.408	5.000	0.000	5.847	3.531	0.912	10.816	0.000	0.800	1.320	0.341	5	0
fried08	5	3.133	4.134	1.849	10.000	0.000	7.375	9.232	4.129	22.450	0.000	0.400	2.074	0.927	4	-1
gaffney11	27	2.309	1.383	0.266	6.333	0.000	10.663	5.942	1.143	30.512	0.000	0.407	0.971	0.187	4	0
garwood	3	5.675	1.217	0.703	6.896	4.462	18.448	3.250	1.877	20.515	14.702	-57.667	50.063	28.904	0	-90
gaudin	8	2.583	3.445	1.218	8.000	0.000	5.460	7.114	2.515	15.362	0.000	-0.500	0.756	0.267	1	-1

gilbert	16	8.002	8.630	2.157	26.000	0.000	16.062	16.582	4.145	57.649	0.000	-2.813	5.332	1.333	10	-8
greenwood	5	6.765	4.692	2.098	12.857	1.500	17.416	11.609	5.191	34.001	5.885	13.000	17.649	7.893	41	0
guanghui	4	4.500	3.786	1.893	10.000	2.000	8.088	4.676	2.338	14.941	5.099	0.250	0.500	0.250	1	0
hill	59	18.499	16.277	2.119	60.300	0.000	79.676	69.999	9.113	299.862	0.000	61.932	84.736	11.032	367	-4
hilton	5	2.333	2.285	1.022	6.000	0.000	5.767	3.678	1.645	8.515	0.000	-0.200	0.837	0.374	1	-1
hospital	5	6.300	3.384	1.513	11.500	4.000	12.313	6.475	2.896	22.130	7.874	0.600	1.342	0.600	3	0
hurley	22	1.496	0.912	0.194	4.452	0.000	5.995	3.454	0.736	16.473	0.000	-5.727	16.977	3.619	34	-24
hutchinson	11	1.431	2.660	0.802	7.385	0.000	5.455	9.627	2.903	27.293	0.000	-5.364	30.303	9.137	76	-33
joyce7	40	7.295	7.071	1.118	38.164	1.750	32.970	23.511	3.717	135.702	10.608	57.075	132.690	20.980	540	-18
kara11	22	8.716	7.155	1.526	26.100	2.000	32.442	27.660	5.897	101.945	8.246	2.909	5.318	1.134	19	0
karasawa	4	7.636	6.917	3.459	16.143	2.000	27.577	15.415	7.707	46.767	15.100	5.000	6.000	3.000	12	0
klug	11	5.742	1.430	0.431	7.617	4.000	15.165	2.311	0.697	18.937	12.374	0.727	11.909	3.591	31	-11
kparrots	11	6.887	4.761	1.435	15.857	1.000	18.233	10.367	3.126	32.413	3.536	39.000	49.520	14.931	132	-4
ksepka	14	4.298	2.366	0.632	8.667	0.000	14.859	7.491	2.002	28.648	0.000	0.429	0.938	0.251	3	0
lambert_seals	9	3.932	2.863	0.954	9.000	0.000	8.859	5.601	1.867	14.906	0.000	4.889	4.702	1.567	14	0
lambert13	21	4.584	4.220	0.921	13.700	1.000	12.715	10.627	2.319	36.816	3.536	2.810	5.354	1.168	16	-3
legg	98	7.380	4.850	0.490	28.667	1.333	52.794	31.540	3.186	173.264	16.147	0.388	1.842	0.186	12	-2
li07	10	2.600	4.018	1.271	14.000	0.667	7.922	4.956	1.567	19.131	5.055	0.600	4.006	1.267	8	-4
lopezRhizo	32	10.115	6.618	1.170	27.403	3.721	27.585	16.823	2.974	66.996	13.308	216.250	684.580	121.018	2838	-240
luo11	59	8.360	7.101	0.924	25.500	1.000	37.783	26.655	3.470	96.703	8.246	14.492	28.352	3.691	126	-4
manos	5	6.871	4.705	2.104	14.000	2.857	16.944	8.129	3.636	28.926	9.215	-10.000	5.385	2.408	-3	-15
mayr05	6	0.306	0.427	0.174	1.000	0.000	1.164	1.779	0.726	4.472	0.000	-9.833	4.215	1.721	-2	-13
mayr11	5	3.100	3.286	1.470	8.500	1.000	8.841	7.688	3.438	20.330	3.317	2.200	0.447	0.200	3	2
mayrea10	4	18.460	13.035	6.518	30.765	3.111	46.074	31.218	15.609	73.053	12.867	895.500	1227.092	613.546	2688	-12
mihalovic	7	1.238	1.007	0.381	2.500	0.000	4.984	4.002	1.513	9.659	0.000	-0.857	2.795	1.056	4	-4
oleary	39	10.195	10.996	1.761	37.000	0.000	45.343	49.768	7.969	168.820	0.000	0.795	1.128	0.181	4	0
perrichot09	7	1.920	0.780	0.295	2.889	0.444	4.694	1.644	0.621	6.118	1.412	-3.714	9.827	3.714	12	-17
pines	40	65.888	25.651	4.056	81.764	0.000	199.127	86.893	13.739	307.920	0.000	1458.150	1914.778	302.753	4999	0
poyato	17	4.177	2.197	0.533	9.793	1.889	11.046	4.294	1.042	21.906	7.038	21.176	81.791	19.837	330	-15
pradel	15	2.513	3.230	0.834	10.800	0.000	6.267	7.221	1.864	21.271	0.000	0.800	1.521	0.393	4	0
prideaux	17	2.094	2.225	0.540	7.600	0.000	11.221	11.124	2.698	39.202	0.000	0.294	0.985	0.239	4	0
puertolas	39	3.273	1.778	0.285	10.500	0.833	16.097	5.776	0.925	37.041	5.949	10.487	48.726	7.802	258	-24
sigwart	26	15.153	9.042	1.773	41.407	4.857	40.054	26.776	5.251	124.875	13.411	64.769	117.070	22.959	425	-12
simm08	5	1.600	1.673	0.748	4.000	0.000	5.875	5.817	2.601	13.126	0.000	1.800	1.643	0.735	3	0
skutchas	11	9.101	8.432	2.542	23.500	0.000	16.210	13.299	4.010	36.889	0.000	12.091	20.447	6.165	53	-5
smith10	8	4.565	8.422	2.978	25.000	0.000	20.645	29.108	10.291	85.157	0.000	0.375	6.232	2.203	15	-4
smith11	6	7.347	13.893	5.672	35.241	0.000	30.602	54.774	22.362	139.595	0.000	7.000	27.408	11.189	60	-21
spaul	35	7.968	6.010	1.016	20.776	1.250	39.135	29.098	4.918	91.198	7.814	44.286	89.303	15.095	497	-1
trueb	12	1.435	1.791	0.517	4.889	0.000	4.085	5.158	1.489	14.397	0.000	2.083	5.282	1.525	16	-1
vea	8	3.663	3.852	1.362	13.177	2.058	15.215	9.599	3.394	38.827	10.221	680.875	588.815	208.178	1550	-1
vlihel	14	1.730	3.390	0.906	12.185	0.000	4.816	6.921	1.850	22.691	0.000	-475.429	113.382	30.303	-240	-596
waterfowl	10	5.224	5.986	1.893	16.085	0.000	18.021	20.195	6.386	51.989	0.000	26.800	145.031	45.863	342	-112
whitlock10	4	3.558	1.064	0.532	5.143	2.889	10.972	2.735	1.367	14.832	8.674	2.750	2.217	1.109	5	0
worthy	11	1.916	1.987	0.599	6.000	0.000	5.999	5.545	1.672	16.108	0.000	0.273	1.849	0.557	4	-2
zhang6	19	2.604	3.187	0.731	9.259	0.000	8.240	9.455	2.169	26.174	0.000	6.895	10.060	2.308	31	-2

Extra Summary Table for Extant Taxa Data

Identifier	Tax	Mean RF	SD RF	SE RF	Max RF	Min RF	Mean PD	SD PD	SE PD	Max PD	Min PD	Mean diff MPTs	SD diff MPTs	SE diff MPTs	Max diff MPTs	Min diff MPTs
adephaga	30	7.045	5.787	1.057	24.051	2.167	25.514	14.276	2.606	60.912	10.908	153.733	421.752	77.001	1664	-12
adnet	16	4.875	4.641	1.160	16.000	0.000	14.127	10.616	2.654	36.056	0.000	2.188	3.834	0.958	14	0
ahyong06	15	4.542	2.906	0.750	12.444	0.000	13.387	6.395	1.651	26.712	0.000	1.133	4.051	1.046	10	-3
apesteg	12	4.500	5.402	1.559	16.000	0.000	8.981	8.328	2.404	24.844	0.000	1.667	2.640	0.762	7	-1
arango7	64	24.975	20.035	2.504	70.400	6.875	96.312	88.748	11.093	280.286	24.409	467.125	850.837	106.355	2968	-16
archostemata	16	6.219	3.167	0.792	13.000	2.000	15.496	6.827	1.707	33.246	6.633	0.688	1.740	0.435	7	0
asher03	28	12.266	15.075	2.849	45.564	0.000	45.862	53.586	10.127	162.240	0.000	64.607	149.796	28.309	724	-4
babot	9	8.994	4.196	1.399	18.000	4.000	21.509	9.498	3.166	41.399	11.662	4.111	3.756	1.252	11	0
beard	4	3.500	5.030	2.515	10.667	0.000	13.952	19.317	9.658	40.958	0.000	-4.000	1.414	0.707	-2	-5
bisconti	12	2.779	1.045	0.302	5.000	1.333	10.424	3.346	0.966	18.163	4.634	-0.500	3.631	1.048	8	-4
blag9	10	1.117	1.863	0.589	5.000	0.000	2.791	4.411	1.395	11.263	0.000	-0.200	1.033	0.327	2	-1
blago4	17	14.310	15.082	3.658	44.800	0.000	45.885	47.384	11.492	149.478	0.000	28.824	68.521	16.619	244	-7
bloch	4	0.500	1.000	0.500	2.000	0.000	1.785	3.571	1.785	7.141	0.000	0.000	0.000	0.000	0	0
boess	5	1.460	0.464	0.207	2.000	0.857	6.366	2.716	1.215	10.569	3.527	-8.400	7.503	3.356	0	-18
bouetel	7	0.371	0.277	0.105	0.714	0.000	1.349	1.003	0.379	2.589	0.000	-9.143	8.859	3.348	0	-20
bourdon	7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0	0
brochu10hornedcrocs	15	7.136	1.951	0.504	12.331	5.519	18.843	4.887	1.262	31.891	15.484	-4.533	264.445	68.279	640	-324
claeson	27	9.242	8.462	1.628	29.925	2.556	33.036	26.077	5.019	99.569	12.003	255.296	680.699	131.001	2304	-15
clarke9	42	6.050	5.338	0.824	19.090	0.667	22.405	14.845	2.291	54.801	5.228	69.857	144.314	22.268	456	0
deng	5	1.257	1.745	0.780	4.000	0.000	3.483	4.675	2.091	10.770	0.000	-21.600	13.428	6.005	-7	-34
dietz	10	3.652	3.049	0.964	7.667	0.000	8.195	6.560	2.074	15.095	0.000	2.900	3.213	1.016	9	0
diogo7	73	4.242	4.357	0.510	19.725	1.125	21.768	16.817	1.968	67.758	6.914	77.370	201.753	23.613	864	-112
diogo8	64	3.700	2.339	0.292	12.712	0.667	18.355	10.482	1.310	52.145	3.859	9.656	22.716	2.839	120	-8
evanidae	5	1.800	1.804	0.807	4.667	0.000	4.495	3.665	1.639	9.774	0.000	0.000	0.000	0.000	0	0
fordyce	7	2.524	0.766	0.290	4.000	2.000	7.200	1.008	0.381	8.944	6.325	0.714	0.756	0.286	2	0
fried08	13	3.388	2.850	0.790	7.429	0.000	7.637	5.819	1.614	15.227	0.000	3.846	7.570	2.100	26	-1
gaffney11	9	2.900	1.347	0.449	5.500	1.000	13.472	5.968	1.989	23.436	4.123	0.778	1.302	0.434	3	-1
garwood	40	8.066	2.306	0.365	15.514	4.688	24.559	3.341	0.528	34.005	18.076	-9.175	47.148	7.455	192	-83
gaudin	8	3.000	2.765	0.977	7.000	0.000	6.315	5.828	2.061	13.982	0.000	0.625	1.188	0.420	2	-1
gilbert	7	23.357	9.949	3.760	28.500	1.000	43.998	18.807	7.109	58.436	3.240	-4.000	4.690	1.773	6	-8
greenwood	21	9.183	5.522	1.205	17.434	1.500	21.079	12.170	2.656	46.658	5.885	8.476	15.683	3.422	49	-2
guanghui	11	3.091	4.425	1.334	13.333	0.000	5.561	6.754	2.036	18.646	0.000	0.455	0.820	0.247	2	0
hill	20	7.499	11.536	2.580	35.632	0.333	32.852	44.493	9.949	135.448	2.068	25.150	52.797	11.806	209	0
hilton	12	3.621	3.065	0.885	9.333	0.000	9.217	7.232	2.088	21.325	0.000	2.667	4.979	1.437	17	0
hospital	16	7.286	4.061	1.015	13.556	2.000	16.371	8.055	2.014	29.643	6.164	2.938	3.991	0.998	12	0
hurley	6	1.805	0.744	0.304	3.000	0.929	6.523	2.007	0.819	9.436	4.080	10.500	13.323	5.439	28	-8
hutchinson	24	2.349	5.117	1.044	19.692	0.000	6.863	12.137	2.477	43.543	0.000	15.250	52.715	10.760	251	-20
joyce7	22	7.312	6.217	1.326	20.333	2.000	36.245	28.383	6.051	105.489	11.356	24.000	67.012	14.287	291	-18
kara11	14	8.369	7.292	1.949	21.000	2.000	36.470	36.022	9.627	108.595	9.950	1.071	1.817	0.486	6	0
karasawa	40	2.292	2.627	0.415	14.840	0.000	12.592	11.104	1.756	52.299	0.000	1.750	7.605	1.202	48	-1
klug	18	7.525	1.521	0.359	10.688	5.750	18.164	2.373	0.559	22.776	15.145	16.889	13.962	3.291	48	0
kparrots	16	3.854	4.493	1.123	14.500	0.667	10.077	8.611	2.153	30.219	2.357	2.813	5.741	1.435	18	-3

ksepka	13	5.769	3.510	0.974	15.333	2.000	18.281	8.816	2.445	43.205	7.071	2.462	6.359	1.764	23	0
lambert13	6	7.079	5.667	2.313	13.714	1.500	19.018	13.836	5.648	36.879	6.036	22.500	39.348	16.064	101	0
lambert_seals	6	0.167	0.408	0.167	1.000	0.000	0.553	1.354	0.553	3.317	0.000	-0.833	0.753	0.307	0	-2
legg	74	8.053	4.707	0.547	27.556	3.333	49.599	27.917	3.245	226.456	27.113	0.162	0.844	0.098	6	0
li07	23	2.186	4.233	0.883	21.000	0.000	7.215	7.565	1.577	34.970	0.000	2.783	6.842	1.427	24	-4
lopezRhizo	6	4.551	0.629	0.257	5.758	3.902	14.389	0.959	0.392	16.234	13.415	0.000	0.000	0.000	0	0
luo11	11	18.960	9.871	2.976	31.778	2.667	71.396	31.215	9.412	99.665	15.500	18.091	19.917	6.005	54	0
manos	21	4.045	1.443	0.315	7.333	1.556	11.040	3.578	0.781	22.880	5.967	2.714	13.020	2.841	35	-15
mayr05	12	1.799	1.969	0.568	5.419	0.000	4.761	4.794	1.384	13.213	0.000	6.750	26.465	7.640	72	-12
mayr11	19	4.229	6.859	1.574	19.273	0.000	9.985	15.240	3.496	45.562	0.000	1.684	3.233	0.742	11	-1
mayrea10	29	18.611	13.443	2.496	33.566	0.286	44.349	29.790	5.532	79.076	1.278	644.207	1225.175	227.509	4979	-18
mihalovic	19	6.226	5.183	1.189	13.783	0.000	16.063	12.769	2.929	39.738	0.000	12.211	17.164	3.938	46	-6
perrichot09	9	2.796	0.947	0.316	4.667	1.733	6.478	1.527	0.509	8.860	4.647	-14.000	6.384	2.128	-2	-22
pines	11	75.571	6.258	1.887	82.254	60.137	229.889	33.181	10.004	273.850	181.537	2001.727	2386.223	719.473	4999	4
poyato	9	7.092	4.558	1.519	15.277	3.333	16.272	9.425	3.142	35.570	9.825	99.333	205.256	68.419	624	0
pradel	3	0.333	0.577	0.333	1.000	0.000	0.943	1.633	0.943	2.828	0.000	0.333	0.577	0.333	1	0
prideaux	17	2.853	2.691	0.653	8.000	0.000	14.713	12.342	2.993	37.108	0.000	0.471	1.007	0.244	3	0
puertolas	11	2.886	0.539	0.163	4.343	2.333	14.426	1.645	0.496	18.609	12.575	4.545	12.136	3.659	40	0
sigwart	7	21.238	9.876	3.733	35.921	10.160	52.821	27.872	10.534	98.741	20.346	119.571	97.872	36.992	274	4
simm08	23	1.510	1.928	0.402	5.200	0.000	5.033	6.184	1.289	16.305	0.000	1.609	2.311	0.482	7	0
skutchas	10	6.313	5.460	1.727	14.647	0.000	12.567	8.702	2.752	23.604	0.000	5.700	12.230	3.867	28	-4
smith10	50	1.835	3.057	0.432	17.333	0.000	9.614	14.681	2.076	76.286	0.000	1.080	6.321	0.894	36	-3
smith11	52	9.910	10.409	1.444	32.141	0.000	44.865	46.427	6.438	132.136	0.000	54.769	115.507	16.018	514	-24
spaul	15	8.394	7.719	1.993	27.796	1.250	33.594	28.931	7.470	96.556	8.376	207.267	699.942	180.724	2734	0
trueb	8	1.146	2.274	0.804	6.500	0.000	3.574	6.723	2.377	19.276	0.000	0.750	2.188	0.773	6	-1
vea	38	3.785	1.982	0.322	8.745	1.082	15.713	6.108	0.991	32.166	5.627	187.526	648.861	105.259	2532	-167
vlihel	14	3.428	3.701	0.989	9.294	0.333	11.621	11.653	3.114	29.687	1.227	-189.571	602.219	160.950	1182	-598
waterfowl	48	8.018	8.902	1.285	35.000	0.000	30.385	31.152	4.496	114.952	0.000	65.875	200.718	28.971	704	-121
whitlock10	22	2.346	1.885	0.402	7.333	0.000	7.429	5.177	1.104	20.478	0.000	0.773	4.859	1.036	15	-4
worthy	14	0.143	0.363	0.097	1.000	0.000	0.655	1.664	0.445	4.583	0.000	-0.214	1.626	0.434	3	-2
zhang6	11	0.766	2.230	0.672	7.429	0.000	2.386	6.113	1.843	19.902	0.000	2.000	7.746	2.335	25	-2
fika	29	15.111	11.370	2.111	50.118	2.327	36.878	27.963	5.193	155.221	10.111	958.966	928.837	172.481	1932	-376
oleary	46	8.401	11.029	1.626	53.333	0.000	36.985	44.349	6.539	167.573	0.000	1.130	1.881	0.277	8	0

Appendix 5.1 The data matrix used for figure 5.1

```
xread 'Data derived from Andres et al 2010 JVP'
34 18
Ornithosuchus_longidens             00000?000?000000000??0000000000000
Herrerasaurus_ischigualastensis     00000000100000000000000000000000000
Scleromochlus_taylori               ??0?0??0?11????0000000?00?00????0?
Preondactylus_buffarinii            0000?01000?100?0??0000010010010000
Eudimorphodon_ranzii                00000010001001?00?0012010011110001
Austriadactytlus_cristatus          0?000010001000?0??????01011011000?
Peteinosaurus_zambelli              ??????????????????0?1?0100??010???
Dimorphodon_macronyx                10000010000100?00?0010000010001000
Campylognathoides_liasicus          00001010001001000100121000000001101
Rhamphorhynchus_muensteri           01001110111001110111111101101221111
Dorygnathus_banthensis              01001110111001100110111101101021111
Scaphognathus_crassirostris         01001010101001000110101001000021110
Sordes_pilosus                      01001010001011000110101000000031110
Jeholopterus_ningchengensis         1?211001?01????01000?0?00000031110
Dendrorhynchoides_curvidentatus     1?21?00??????????1????0?0000?031???
Batrachognathus_volans              1?211001?01????0100010?00000031110
Anurognathus_ammoni                 1?21100???1?1?00100010?00000031110
Pterodactyloidea                    01001011?10?110101111010000031100
;
proc/;
```

Appendix 5.2 Script to implement selective-permutations ' mher.run '

```
macro= ;

resettime ;

if ( argnumber )

   if ( eqstring [ %1 start ] )

      macro - ; /* turn off macro in order to setup enough memory below */

      macro *10 (1000 + (4*root) + ( root*(nchar+1) )) ;

      macro [ 60000 ;

      macro=;



      var:
```

```
         started
         matrix [ root (nchar+1) ] i ; /* store in memory current matrix */
      set started 773 * nchar * ntax ;
      loop 0 ntax
         loop 0 nchar
            set matrix[ #1 #2] states[ #2 #1] ;
            stop
         stop
      set i time ;
      quote Initialization took '/.0i' sec.;
      proc/;
      end
   end
var:
  started matrix [ root (nchar+1) ] i j k
  nonmiss[ root ] rlist [ root ] cur seen reverse[ root ] ;
if ( 'started' != ( 773 * nchar * ntax ) )
   errmsg You havent initialized!!;  /* Sincere thanks to Pablo Goloboff */
   end
report - ;
loop 0 nchar
   progress #1 nchar Scrambling ;
   set j 0 ;
   loop 0 ntax
```

```
      if ( states[ #1 #2] == missing ) continue ; end
      set nonmiss[ 'j' ] #2 ;
      set reverse [ #2 ] 'j' ;
      set j ++ ;
      stop
   if ( !'j' ) continue ; end  /* if matrix has no non-missing data */

   set rlist randomlist [ 'j' ] ;
   set j -- ;

   loop 0 'j'
      set cur 'matrix [ 'nonmiss [ 'rlist [ #2 ]' ]' #1 ]' ;
      xread =! #1
         'nonmiss [ #2 ]' $bitset 'cur' ;
      stop
   stop
progress/;

set j time ;
quote Permuting took '/.0j' sec.;

report= ;
xread == ;
proc/;
```

Appendix 5.3 Script to calculate MHER in one command 'get-mher.sh'

Pre-requisites:

- a unix operating system (Linux/Mac),
- a reasonably new version of TNT installed that can be called from the terminal by typing 'tnt'. The script relies on it being called tnt, so if it is named anything different it will NOT work. 'mher.run' utilizes some of the newer features in TNT so it will NOT work with older pre-2011 versions of TNT.

Instructions for use:

- place 'get-mher.sh' , 'reps.txt' , and 'mher.run' in the same directory as the .tnt formatted cladistic data matrix you would like to test.
- Make sure your .tnt data file ends with `procedure /;` Anything else like `proc /;` or `p /;` despite being 'valid' TNT shorthand will NOT work with this get-mher.sh bash script.
- 'reps.txt' contains instructions to perform 1000 selective-permutations, which are needed to calculate the modified-MEANNS. I have deliberately hardcoded-in 1000 replications, to prevent people from being tempted to perform a statistically insufficient number of replications e.g. 10 or 100. However, if you do want to change the number of replications to more or less, this is the file that needs to be changed.
- For ease-of-use and forking, this bundle of files ('reps.txt' , 'mher.run' & 'get-mher.sh') used to calculate MHER are provided online on github at:
  https://github.com/rossmounce/thesis_ESM/tree/master/MHER

The bash script itself including license notice and comments:

```
#!/bin/bash

#This is a bash script to perform the modified Homoplasy Excess Ratio
#on a dataset of your choice passed to this script as an argument
#You also need the files 'mher.run' AND 'reps.txt' in the same dir

#The MIT License (MIT)

#Copyright (C) 2013 Ross Mounce

#Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
the Software, and to permit persons to whom the Software is furnished to do so,
subject to the following conditions:

#The above copyright notice and this permission notice shall be included in all
copies or substantial portions of the Software.

#THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
```

```
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

#DATA FILE TO BE TESTED MUST END with 'procedure /;'
#this makes a copy of the data file with
#additional instructions appended to tmp.tnt
sed 's@procedure \/;@log temp1.log; minmax\*; sect:slack 40; xmult=level10;
log\/; log output_reps.log; mher start; proc reps.txt; quit;@' $1 > tmp.tnt

#ensure that all inapplicables have been converted to ? marks
sed -i 's/-/?/g' tmp.tnt;
#put back hyphens in ccode block if they were taken out by the above command

sed -i 's/?\[\//-\[\//g' tmp.tnt;

#this will do everything then quit. output hardcoded to output_reps.log
#this step may take a LONG time
#perhaps hours depending on the size of your data
tnt proc tmp.tnt;

printf "MEANNS calculations in TNT are complete \n"

#hacky shorter way of gettings the modified-MEANNS
grep 'Best score:' output_reps.log | awk '{sum+=$3} END { print sum/NR}' >
mns.tmp

#get MINL, temp1 is the first logfile output from TNT
head -1 temp1.log | sed 's/\// /g' | cut -d ' ' -f 8 > minl.tmp
#get L, L=tmp3.tmp, MINL=tmp2.tmp
tail -1 temp1.log | sed 's/\./ /g' | cut -d ' ' -f 3 > l.tmp

#print results
printf "MINL = `cat minl.tmp` \n"
printf "L = `cat l.tmp` \n"
printf "Modified-MEANNS = `cat mns.tmp` \n"
paste mns.tmp minl.tmp l.tmp | awk '{o = ($1-$3)/($1-$2)} END { print "Modified-
HER = " o }'

#clean up temporary files but leave behind permuted matrix log file
rm temp1.log;
rm tmp.tnt;
rm minl.tmp;
rm l.tmp;
rm mns.tmp;
```

Appendix 6.1 Demonstrating grep

Sample output from the command

find . -name ".txt" -print0 | xargs -0 -n 1000 -P 3 grep -iR -m1 -A 4 -B 4 "paup[^heary]"*

as applied to the Zootaxa corpus. The four lines of additional context above and below make it very easy to classify the usage of PAUP* with certainty in 99% of cases as either in the context of a molecular study, or a morphological study. The first two given here are clearly molecular whilst the third is clearly morphology-based.

./2011/zt02768p031.txt-Phylogenetic analysis. Divergence and polymorphism in cox1 sequences

./2011/zt02768p031.txt-Cox1 sequences from complete mitochondrial genomes of Halisarca harmelini (this study) Halisarca dujardini

./2011/zt02768p031.txt-from White Sea (NC_010212) and Chondrilla aff. nucula (NC_010208) were aligned with partial cox1 sequences

./2011/zt02768p031.txt-of Halisarca dujardini from North Sea (this study) and Chondrilla nucula (Duran & Rützler 2006) using ClustalW

./2011/zt02768p031.txt:2.0.11 (Larkin et al. 2007). The alignment was trimmed manually to remove terminal gaps associated with incomplete sequences. **PAUP***v.4b10 was used to build a neighbor-joining tree based on uncorrected “p” distances.

./2011/zt02768p031.txt-Phylogenetic analysis of demosponge relationships. Amino-acid sequences for Cantharellus cibarius

./2011/zt02768p031.txt-mtDNA were downloaded from http://megasun.bch.umontreal.ca/People/lang/FMGP/proteins.html; those for Capsaspora owczarzaki mtDNA were provided by Franz Lang (Université de Montréal). Other sequences were derived

./2011/zt02768p031.txt-from the GenBank files: Agelas schmidti EU237475, Amphimedon compressa NC_010201, Amphimedon queenslandica NC_008944, Aplysina fulva NC_010203, Axinella corrugata NC_006894, Chondrilla aff. nucula

./2011/zt02768p031.txt-NC_010208, Callyspongia plicifera NC_010206, Cinachyrella kuekenthali EU237479, Ectyoplasia ferox

--

./2011/zt02767p040.txt-observed among the O. ishikawae taxa. Gap sites between the 16S units of O. ishikawae and the other taxa were

./2011/zt02767p040.txt-treated as missing data in the following analyses. Uncorrected p values (nucleotide changes per compared sequence

./2011/zt02767p040.txt-length) between taxa were calculated from the resultant alignments. The phylogeny was analyzed by the maximum

./2011/zt02767p040.txt-likelihood (ML) and maximum parsimony (MP) methods. ML and MP analyses were performed using

./2011/zt02767p040.txt:TREEFINDER ver. Oct. 2008 (Jobb 2008) and **PAUP** 4.10b (Swofford 2003), respectively. For the ML analysis,

./2011/zt02767p040.txt-we applied the J2 (Rodriguez et al. 1990) + gamma (8 categories and shape parameter = 0.32) substitution model,

./2011/zt02767p040.txt-which was estimated using Akaike’s information criterion (AIC) implemented in KAKUSAN3 software (Tanabe

./2011/zt02767p040.txt-2007). The robustness of the resultant ML and MP trees were evaluated using bootstrap probabilities calculated

./2011/zt02767p040.txt-from nonparametric bootstrap analyses with 500 pseudoreplications.

--

./2011/zt02923p047.txt-The character matrix was edited (Table 2) employing the software Nexus Data Editor

v0.5.0 (Page 2001a). All

./2011/zt02923p047.txt-characters (35 binary and 5 multistate) were set as unordered and equally weighted; the multistate characters were

./2011/zt02923p047.txt-interpreted as “uncertainty”, and the gaps were treated as “missing”. Trees were rooted by the outgroup method.

./2011/zt02923p047.txt:We carried out a parsimony analysis (Exhaustive Search) in **PAUP** 4.0b.10 (Swofford 1998–2002), using the

./2011/zt02923p047.txt-default settings of the software. The MaxTrees limit was set to automatically increase from the initial setting. The

./2011/zt02923p047.txt-resulting trees were examined with TreeView 1.6.6 (Page 2001b) and TreeGraph2 (Stöver & Müller 2010).

./2011/zt02923p047.txt-

./2011/zt02923p047.txt-30 · Zootaxa 2923 2011 Magnolia Press

Appendix 6.2 Demonstrating search queries

Search terms and corresponding URLs used to generate the results presented in Table 4:

MAS: PLOS ONE, 2006-2009, phylip

http://academic.research.microsoft.com/Detail?searchtype=4&query=year%3E%3d2006%20year%3C%3d2009%20jour%3a%28plos%20one%29%20phylip

MS: Zootaxa, (all years), phylogeny

[hits published later than 2013-07-11, outside the corpus scope were manually removed from the count after the search results were returned]

http://www.mendeley.com/research-papers/search/?query=phylogeny+AND+published_in%3AZootaxa

WoK: Topic=(phylogen*) AND Publication Name=(plos one) Timespan=2006-2009. Search language=Auto

http://apps.webofknowledge.com/summary.do?SID=T2FmvrZ1heYnZ9rNrlG&product=UA&qid=1&search_mode=GeneralSearch

Scopus: query: ALL(phylogeny) AND SRCTITLE(plos one) AND PUBYEAR > 2005 AND PUBYEAR < 2010

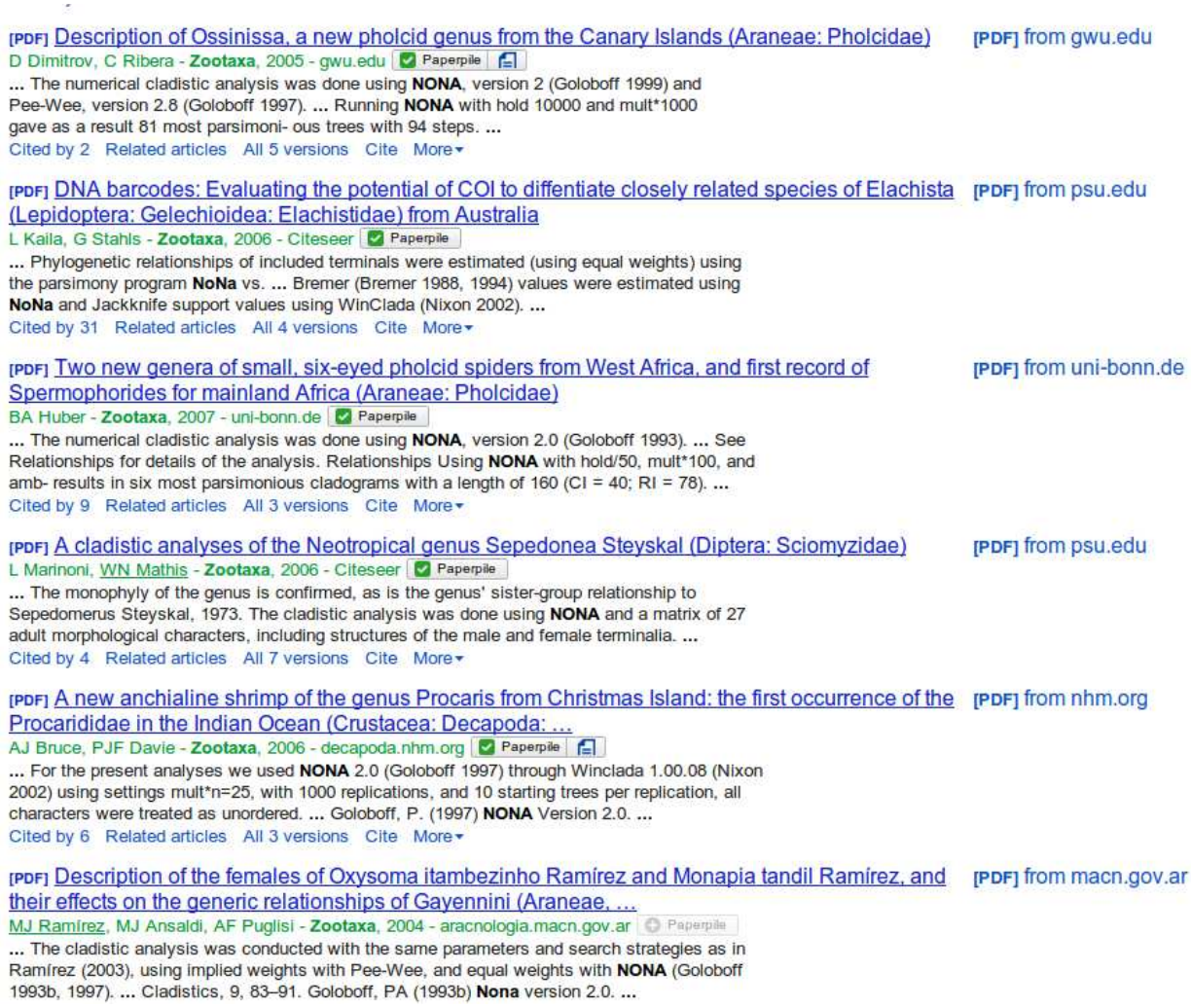
http://www.scopus.com/results/results.url?sort=plf-f&src=s&st1=phylogeny&searchTerms=PLOS+ONE%3f%21%22*%24&sid=7053A17CC766D7A14AC3E36512AC1E40.ZmAySxCHlBxxTXbnsoe5w%3a80&sot=b&sdt=b&sl=75&s=ALL%28phylogeny%29+AND+SRCTITLE%28PLOS+ONE%29+AND+PUBYEAR+%3E+2005+AND+PUBYEAR+%3C+2010&origin=searchbasic&txGid=7053A17CC766D7A14AC3E36512AC1E40.ZmAySxCHlBxxTXbnsoe5w%3a8

GS: with all of the words: phylogeny , Return articles published in: PLOS ONE, Return articles dated between: 2006 - 2009

http://scholar.google.co.uk/scholar?as_q=phylogeny&as_epq=&as_oq=&as_eq=&as_occt=any&as_sauthors=&as_publication=PLOS+ONE&as_ylo=2006&as_yhi=2009&btnG=&hl=en&as_sdt=0%2C5

Appendix 6.3 Google Scholar finds IR copies

Screenshot of the search results returned by GS for the 'nona' search which anecdotally provide evidence that GS generally only finds terms in the full-text of the paper if the full-text of the paper has been optionally and additionally deposited by the authors on 'academic web addresses' that are crawled by Google. In this example we can see many examples of this e.g. Huber's 2007 Zootaxa paper which is freely available from the University of Bonn repository here: http://www.uni-bonn.de/~bhuber1/PDFs/Huber_2007_Anansus_Nyikoa.pdf



[PDF] Description of Ossinissa, a new pholcid genus from the Canary Islands (Araneae: Pholcidae) [PDF] from gwu.edu
D Dimitrov, C Ribera - **Zootaxa**, 2005 - gwu.edu Paperpile
... The numerical cladistic analysis was done using **NONA**, version 2 (Goloboff 1999) and Pee-Wee, version 2.8 (Goloboff 1997). ... Running **NONA** with hold 10000 and mult*1000 gave as a result 81 most parsimoni- ous trees with 94 steps. ...
Cited by 2 Related articles All 5 versions Cite More

[PDF] DNA barcodes: Evaluating the potential of COI to diffentiate closely related species of Elachista (Lepidoptera: Gelechioidea: Elachistidae) from Australia [PDF] from psu.edu
L Kaila, G Stahls - **Zootaxa**, 2006 - Citeseer Paperpile
... Phylogenetic relationships of included terminals were estimated (using equal weights) using the parsimony program **NoNa** vs. ... Bremer (Bremer 1988, 1994) values were estimated using **NoNa** and Jackknife support values using WinClada (Nixon 2002). ...
Cited by 31 Related articles All 4 versions Cite More

[PDF] Two new genera of small, six-eyed pholcid spiders from West Africa, and first record of Spermophorides for mainland Africa (Araneae: Pholcidae) [PDF] from uni-bonn.de
BA Huber - **Zootaxa**, 2007 - uni-bonn.de Paperpile
... The numerical cladistic analysis was done using **NONA**, version 2.0 (Goloboff 1993). ... See Relationships for details of the analysis. Relationships Using **NONA** with hold/50, mult*100, and amb- results in six most parsimonious cladograms with a length of 160 (CI = 40; RI = 78). ...
Cited by 9 Related articles All 3 versions Cite More

[PDF] A cladistic analyses of the Neotropical genus Sepedonea Steyskal (Diptera: Sciomyzidae) [PDF] from psu.edu
L Marinoni, WN Mathis - **Zootaxa**, 2006 - Citeseer Paperpile
... The monophyly of the genus is confirmed, as is the genus' sister-group relationship to Sepedomerus Steyskal, 1973. The cladistic analysis was done using **NONA** and a matrix of 27 adult morphological characters, including structures of the male and female terminalia. ...
Cited by 4 Related articles All 7 versions Cite More

[PDF] A new anchialine shrimp of the genus Procaris from Christmas Island: the first occurrence of the Procaridídae in the Indian Ocean (Crustacea: Decapoda: ... [PDF] from nhm.org
AJ Bruce, PJF Davie - **Zootaxa**, 2006 - decapoda.nhm.org Paperpile
... For the present analyses we used **NONA** 2.0 (Goloboff 1997) through Winclada 1.00.08 (Nixon 2002) using settings mult*n=25, with 1000 replications, and 10 starting trees per replication, all characters were treated as unordered. ... Goloboff, P. (1997) **NONA** Version 2.0. ...
Cited by 6 Related articles All 3 versions Cite More

[PDF] Description of the females of Oxysoma itambezinho Ramírez and Monapia tandil Ramírez, and their effects on the generic relationships of Gayennini (Araneae, ... [PDF] from macn.gov.ar
MJ Ramírez, MJ Ansaldi, AF Puglisi - **Zootaxa**, 2004 - aracnologia.macn.gov.ar Paperpile
... The cladistic analysis was conducted with the same parameters and search strategies as in Ramírez (2003), using implied weights with Pee-Wee, and equal weights with **NONA** (Goloboff 1993b, 1997). ... Cladistics, 9, 83–91. Goloboff, PA (1993b) **Nona** version 2.0. ...

Appendix 7.1 Phylogenetic position of Diania challenged

This section was co-authored with Matthew A. Wills and published (Mounce & Wills, 2011)

Formatting has been changed to avoid copyright infringement with the published version

Liu *et al.* (2011a) describe a new and remarkable fossil, *Diania cactiformis*. This animal apparently combined the soft trunk of lobopodians (a group including the extant velvet worms in addition to many Palaeozoic genera) with the jointed limbs that typify arthropods.

They go on to promote *Diania* as the immediate sister group to the arthropods, and conjecture that sclerotized and jointed limbs may therefore have evolved before articulated trunk tergites in the immediate arthropod stem. The data published by Liu *et al.* (2011a) do not unambiguously support these conclusions; rather, we believe that *Diania* probably belongs within an unresolved clade or paraphyletic grade of lobopodians. Without taking issue with the interpretation of *Diania* offered by Liu *et al.* (2011a), or of the manner in which they coded their characters, we were nonetheless unable to derive their cladogram optimally from the data published. Moreover, we could not replicate their results using any other plausible optimality criteria, or by varying additional parameters not specified by the authors. Liu *et al.* (2011a) report analysing their data in PAUP* under maximum parsimony and with implied weights using k = 2 (a rather arbitrary choice), but do not mention any other assumptions (for example, the imposition of character order). They obtained three most parsimonious trees, each of 130 steps. Straightforward replication of their stated settings yields 13 trees of just 90 steps each, the strict consensus of which is illustrated (Fig. 7.1).

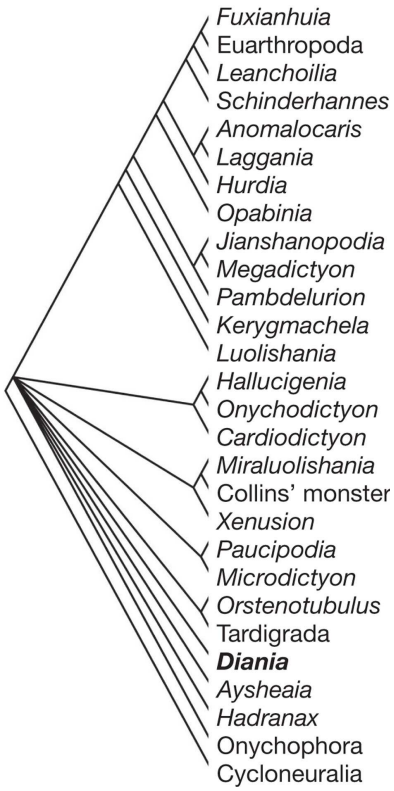


Figure 7.1 The strict consensus of 13 most parsimonious trees (L=90) obtained from the published data and settings specified by the authors.

Why such a difference? Several of their characters contained inapplicable or gap codings.

These appear where a ‘daughter’ character is logically contingent upon the state of a ‘parent’, and cannot be coded when the parent is absent. For example, character 6 (position of frontal appendage) can only be coded in taxa that possess a frontal appendage (character 5) in the first instance (such that a ‘‘0’’ for character 5 necessitates a ‘‘-’’ for character 6). In morphological analyses such as this, inapplicable states are usually assumed to have no bearing on the analysis, being reconstructed passively in the light of known states. In analyses of nucleotide data, by contrast, gaps may alternatively be construed as a fifth and novel state, because shared deletions from some ancestral sequence may actually be informative. If this assumption is made with morphological data, however, all the logically uncodable states in a character are initially assumed to be homologous, and a legitimate basis for recognizing clades. At best, this assigns double weight a priori to absences in the ‘parent’ character (because the daughter is always contingent), and at worst is positively misleading. This is the approach that we believe Liu *et al.* (2011a) may have taken. Reanalysis of their data using ‘gapmode = newstate’ combined with ‘collapse 5= MinBrLen’ settings in PAUP* produced some optimal trees of 130 steps. However, we were still unable to replicate the relationships shown in their Fig. 4, even when varying k between 0 and 10. Rather we either resolved *Diania* in a basal polytomy, or slightly higher in the tree but separated from the arthropods by at least five nodes. At best, therefore, the position of *Diania* is highly labile and extremely sensitive to the precise methods used. We certainly feel that it is premature to draw conclusions regarding its supposedly pivotal position in the evolution of arthropods. However, our reanalyses do not challenge the more general conclusions of Liu *et al.* (2011a): namely that the full complement of arthropod characters were probably acquired piecemeal and possibly convergently. Many closely allied groups exploited successfully some but not all of the characters that typify the arthropod crown group. Only in retrospect do we discern a single, ladder-like trajectory through what was really a much more eccentrically branching bush.